Experiments on non-orthogonal signal representation

Evangelos Sagianos

MSc by Research in Pattern Analysis and Neural Networks

ASTON UNIVERSITY

September 2003

ASTON UNIVERSITY

Experiments on non-orthogonal signal representation

Evangelos Sagianos

MSc by Research in Pattern Analysis and Neural Networks, 2003

Thesis Summary

The problem of non-orthogonal signal representation we consider here consists of expressing the signal as a linear superposition of non-orthogonal waveforms. Such waveforms are selected from a set of functions, in general redundant, which is called a dictionary. The techniques we discuss to address the problem are in the line of the adaptive matching pursuit methods. Potential advantages, arising from the freedom to decide on the dictionary, are illustrated by considering dictionaries of different nature. Further potential advantages, arising simply by relaxing the orthogonality condition, are illustrated by introducing a non-orthogonal set of wavelets which originates from the orthogonal Haar basis.

Keywords: non-orthogonal signal representation, adaptive techniques, matching pursuit techniques, greedy algorithms, non-orthogonal Haar wavelets, biorthogonal sets, sparse representations, atomic decomposition.

Acknowledgements

For this project I would like to thank most of all, Dr. Laura Rebollo-Neira. Thanks to her efforts and guidance I have managed to complete this work.
I would like to thank Prof A. Constantinides and Prof D. Saad, for evaluating the thesis.
I would also like to thank my parents, who gave me the opportunity to study here, and the NCRG teachers and staff, for creating a wonderful environment for studying.
Many thanks to Benoit F. , Benoit M. , Fred, Hiroko, Jodie and Remi for the help and support throughout this year, and most important for their friendship.
Some other people I would like to thank for making it a great year in Aston University, are Nikos Skantzos, Amos Martinez, Kate, Becky, Ahad, Lucy Luca, Petros, Christina, Panos, Roberto, Mandy and Chiara.

I know nothing, except the fact of my ignorance. *Socrates*

Contents

Chapter 1

Introduction

Traditional methods for signal representation involve the use of orthogonal bases, such as the Fourier basis or the later introduced orthogonal wavelet bases [3]. The problem of N-term approximation, where one fixes a basis and looks to approximate a signal by a linear combination of N-terms of the basis, is a nonlinear problem. The nonlinearity is a consequence of allowing the terms to depend on the signal being approximated. The problem of finding the best of such N-term approximations has been, and still is, subject of much mathematical work even in the case of orthogonal basis.

More recently, there has emerged another more complicated form of nonlinear approximation, which we can call a highly nonlinear approximation. It takes many forms but has one basic ingredient: a basis is replaced by a system of functions which is usually redundant. Some types of approximations that fall into this general category are mathematical frames [4, 21], adaptive pursuit [8, 6, 11, 16, 13] and adaptive basis selection [1, 2]. This relatively new setting for signal representation seems to offer much promise for greater effectiveness in terms of approximation rate and sparseness. On the other hand it gives rise to highly nontrivial theoretical and practical problems. It can be said that a rigorous theory is only now emerging, and certainly far beyond the scope of this work.

The purpose of this project was to produce evidences of the potential advantages of highly nonlinear signal representation outside the basis setting. With this motivation, and also with the ambition of drawing, somehow, theoretical conclusions, we have designed a series of numerical experiments.

The methods that we have considered to deal with the problem of nonorthogonal signal approximation are all in the line of adaptive pursuit. We present a numerical example with the aim of comparing these approaches with regard to convergence rate. However, our central aim was not to focus on comparing methodologies. We have been lead for a different motivation. On the one hand our goal was to illustrate by some examples what is a rather obvious remark:

Relaxing the orthogonality condition gives us more freedom to choose the spanning

set. Therefore, we can obtain higher quality approximations of a given signal by using the adequate spanning set.

On the other hand we have been driven by the motivation of answering the following question:

Is it possible to obtain a significant gain in the approximation problem only by relaxing the condition of orthogonality?

In order to answer this question we have proposed a generalisation of the Haar wavelet system. Such a system is an orthogonal basis if the translation parameter is an integer. By relaxing this condition on the translation parameter one creates an overcomplete set. We show here that, in the finite dimension case, introducing a non integer translation parameter not only introduces redundancy but also enlarges the corresponding subspace. We believe this to be the most important contribution of this work.

The thesis is organised as follows: In chapter 2 we describe adaptive methods for non-orthogonal signal representation which are known as adaptive pursuit techniques. The above mentioned non-orthogonal Haar system is proposed in chapter 3 where a series of experiments leading to definitive theoretical conclusions is presented. In chapter 4 we illustrate, by very simple examples, the advantage of having the freedom of choosing different types of waveforms for approximating signals of different nature. Partial conclusions are given in each chapter and some general conclusions are drawn in Chapter 5.

Chapter 2

Adaptive techniques for signal representation

2.1 Introduction

In this section we discuss the problem of non-orthogonal signal representation, which is often referred to as atomic decomposition. We also comment on some essential differences between orthogonal and non-orthogonal signal approximations. In the following sections we present some techniques for adaptive signal representation. All the techniques we consider are leading to the so called adaptive pursuit approach.

2.1.1 Signal space

As our signal space we adopt the Hilbert space $\mathcal{H}$ of square-integrable functions. Hence, a signal f belongs to $\mathcal{H}$ if

$$\| f \| = \sqrt{\int |f(x)|^2 dx} < \infty.$$

The inner product of two functions $f, g \in \mathcal{H}$ is defined as in [18]:

$$\langle f, g \rangle = \int f^*(x) g(x) dx,$$

where with f^* we denote the complex conjugate of f.

2.1.2 Atomic decomposition

Recent methodologies for signal representation operate on decomposing an arbitrary signal to a linear expansion of waveforms [9, 7]. Such a representation is known as *atomic decomposition*. We give next the formal definition.

Definition: The atomic decomposition of a signal $f \in \mathcal{H}$ is its representation as a linear superposition

$$f = \sum_{n=1}^{N} c_n \alpha_n. \tag{2.1}$$

The waveforms $\alpha_n \in \mathcal{H}$ are called *atoms* and are chosen, by some criteria, from a redundant family of functions, which is called a *dictionary.*

Definition: Let the family of atoms $\mathcal{D} = \{\alpha_n\}_{n \in \mathbb{Z}}$, be a dictionary for $\mathcal{H}$. If this dictionary contains a dictionary of linearly independent atoms, the later is a basis for $\mathcal{H}$ [1].

According to the atoms in $\mathcal{D}$, the dictionary can be described as *orthogonal* or *non-orthogonal.* An orthonormal set of atoms forms an orthogonal dictionary, whereas non-orthogonal dictionaries include bases and overcomplete sets. Thus, non-orthogonal dictionaries can be constructed from all sorts of waveforms.

Given a signal, the actual computation of its representation as given in (2.1), depends very much on the nature of the dictionary. In the next section we discuss a major difference arising in iterative signal approximation techniques, when dealing with orthogonal or non-orthogonal dictionaries.

2.1.3 Orthogonal and non-orthogonal signal representation

Let us assume that $f \in \mathcal{H}$ and the atoms $\alpha_n \in \mathcal{H}$ involved in (2.1) are given. Let us also assume that these atoms form an orthonormal set. Hence, the coefficients c_n in (2.1) can be obtained in a straightforward manner, as the inner products:

$$c_n = \langle \alpha_n, f \rangle \;\; ; \;\; n = 1, \; \ldots \; , N. \tag{2.2}$$

If the signal we are representing does not belong to the subspace spanned by the N orthogonal atoms α_n, the coefficients given in (2.2) are guaranteed to provide the optimal approximation of the signal in the corresponding subspace. (The approximation is optimal in a minimum distance sense [10]). Moreover, if in order to improve the approximation, we add one atom to the linear expansion, the new optimal approximation is obtained as:

$$f = \sum_{n=1}^{N} c_n \alpha_n + c_{N+1} \alpha_{N+1}, \tag{2.3}$$

$$\text{with} \;\; c_{N+1} = \langle \alpha_{N+1}, f \rangle. \tag{2.4}$$

Equivalently, if an atom is removed, say the atom α_j, the optimal linear expansion in the reduced subspace is given by:

$$f = \sum_{\substack{n=1 \\ n \neq j}}^{N} c_n \alpha_n - c_j \alpha_j, \tag{2.5}$$

$$\text{with} \quad c_j = \langle \alpha_j, f \rangle. \tag{2.6}$$

When relaxing the orthogonality condition the process of iterative signal approximation becomes far more complicated than the simple procedure described above. This is a consequence of the fact that, in this case, if the representation subspace is increased (or reduced), the coefficients of the corresponding linear expansion should be appropriately modified for them to yield an optimal representation in the new subspace [16, 13]. To be precise: Let us consider that the atoms α_n are not orthogonal, and let $\sum_{n=1}^{N} c_n^{(N)} \alpha_n$ be the best approximation of a given signal in the span of the N atoms α_n. The superscript of the coefficients indicates the dependence of these coefficients on the number of atoms being considered. Hence, if the number of atoms is increased as above, by considering one more atom α_{N+1}, the corresponding optimal approximation of the signal is to be computed as

$$f = \sum_{n=1}^{N} c_n^{(N+1)} \alpha_n + c_{N+1}^{(N+1)} \alpha_{N+1}. \tag{2.7}$$

Unlike in the orthogonal case (cf (2.3)) now **all** coefficients must be recalculated for them to give rise to an optimal approximation.
Equivalently, when removing an atom, say the j-one, in order to obtain an optimal approximation the remaining coefficients should be recalculated, i.e. the new approximation is to be computed as

$$f = \sum_{\substack{n=1 \\ n \neq j}}^{N} c_n^{(N-1)} \alpha_n - c_j^{(N-1)} \alpha_j. \tag{2.8}$$

The need for recalculating coefficients, when using non-orthogonal atoms in adaptive approximation of a signal, entails a practical complication one has to face. Here, we will address the problem adopting a recently introduced biorthogonalisation technique. Such a technique allows for recursive modification of the coefficients so as to achieve an approximation which is optimal in a minimum distance sense [16, 13].
Another problem that we have to address when approximating a signal by using non-orthogonal atoms, is the one of deciding how to choose the atoms to be used for the signal approximation. This is a very complex problem and several solutions have been proposed with different purposes in mind [8, 2, 1]. Here we shall restrict our considerations to methods that have been developed along the line of the so called

Matching Pursuit methodology [8, 6, 13, 11, 16]. In the following sections we present a description of such methodologies dedicating a special attention to the Optimised Orthogonal Matching Pursuit approach, since this is the technique that we will use in our experiments.

2.2 The Matching Pursuit (MP) Approach

Let $\mathcal{D} = \{\alpha_n; n = 1, \cdots, N\}$ be a finite dictionary and $f \in \mathcal{H}$ a given signal to be represented as a linear superposition of atoms chosen from $\mathcal{D}$. The Matching Pursuit (MP) approach introduced in [8] proposes to make the selection by successive approximations of f. At each step the corresponding residue is sub-decomposed by projecting it onto the dictionary atom that matches it best.

Let R_k be the k-th order residue and l_k the index n for which the corresponding dictionary atom α_{l_k} yields a maximal value of $|\langle \alpha_n, R_k \rangle|$; $n = 1, \ldots, N$, i.e.

$$\alpha_{l_k} = \text{argmax}_{\alpha_n} |\langle \alpha_n, R_k \rangle|. \tag{2.9}$$

Starting with $f_1 = 0$ and $R_1 = f$, the $k-$th order residual is decomposed to:

$$R_k = \langle \alpha_n, R_k \rangle \alpha_n + R_{k+1}, \tag{2.10}$$

which defines the residual of order (k+1).
Since R_{k+1} given in (2.10) is orthogonal to α_n we have

$$||R_k||^2 = |\langle \alpha_n, R_k \rangle|^2 + ||R_{k+1}||^2. \tag{2.11}$$

From the above equation it follows that the dictionary atom α_{l_k} yielding a maximal value of $|\langle \alpha_n, R_k \rangle|$ minimises $||R_{k+1}||^2$.
Hence, in order to minimise the residual error $||R_{k+1}||^2$, we have to maximise the value of $|\langle \alpha_n, R_k \rangle|$.

Letting the algorithm evolve with k, equation (2.10) gives the representation of the signal as:

$$f = f_k + R_{k+1}, \tag{2.12}$$

where f_k is:

$$f_k = \sum_{n=1}^{k} \langle \alpha_{l_n}, R_n \rangle \alpha_{l_n}. \tag{2.13}$$

The above described algorithm operates through the following steps:

Initialise $k = 1$, $R_1 = f$, $f_1 = 0$ and set some tolerance parameter $\epsilon > 0$ for the residual error.

Step 1: For $n = 1, \ldots, N$, compute $|\langle \alpha_n, R_k \rangle|$ and choose the atom giving the maximal value. Denote such an atom α_{l_k}.

Step 2: Compute R_{k+1} by the formula: $R_{k+1} = R_k - \langle \alpha_n, R_k \rangle \alpha_n$.

Step 3: If the residual $||R_{k+1}||^2 \leq \epsilon$ then stop the algorithm, else set $k = k+1$ and repeat steps $1 \rightarrow 3$.

The MP approach is a fast technique which, as it is clear from the operational steps, is very easy to implement. However, although its asymptotic convergence has been proved [7], in some situations the convergence may be so slow that it is impossible to achieve in practice. A refinement of this technique, which is based on orthogonalisation of the atoms, improves the convergence rates and provides convergence in a finite number of steps. This technique is described in the next section.

2.3 Orthogonal Matching Pursuit (OMP)

The MP approximations are improved by orthogonalising the directions of projection, as proposed in [11, 8]. The resulting orthogonal pursuit converges with a finite number of iterations, which is not the case for a non-orthogonal pursuit.

The atom $\alpha_{l_{k+1}}$ selected by the MP is a priori not orthogonal to the previously selected atoms $\{\alpha_{l_n} \; ; \; n = 1, \ldots, k\}$. When subtracting the projection of the residual over $\alpha_{l_{k+1}}$ the algorithm reintroduces new components in the directions of $\{\alpha_{l_n} \; ; \; n = 1, \ldots, k\}$. This is avoided by projecting the residues on an orthogonal family $\{\phi_n \, ; \, n = 1, \ldots, k\}$ computed from $\{\alpha_{l_n} \, ; \, n = 1, \ldots, k\}$.

Let us initialise $\phi_1 = \alpha_{l_1}$, and $R_1 = f$. The atom α_{l_1} is selected so as to maximise $|\langle \alpha_n, R_1 \rangle|$. The OMP evolves for $k \geq 2$, by selecting the atom α_{l_k} that maximises the expression $|\langle \alpha_n, R_k \rangle|$, i.e.

$$\alpha_{l_k} = \text{argmax}_{\alpha_n} |\langle \alpha_n, R_k \rangle|. \tag{2.14}$$

The orthogonal function ϕ_k is obtained as follows:

$$\phi_k = \alpha_{l_k} - \sum_{n=1}^{k-1} \frac{\langle \alpha_{l_k}, \phi_n \rangle}{\| \phi_n \|^2} \phi_n, \tag{2.15}$$

and the signal is expanded as:

$$\begin{aligned} f &= \sum_{n=1}^{k} \frac{\langle R_n, \phi_n \rangle}{\| \phi_n \|^2} \phi_n + R_{k+1} \\ &= \hat{P}_{V_k} f + R_{k+1}, \end{aligned} \tag{2.16}$$

where $\hat{P}_{V_k}$ is the orthogonal projector on the space V_k generated by $\{\phi_n\}_{1 \leq n \leq k}$, and R_{k+1} is the residual of the next order.

For any $k \geq 1$ the residue R_k is the component of f orthogonal to V_{k+1}. This implies that:

$$\langle R_k, \phi_k \rangle = \langle R_k, \alpha_{l_k} \rangle. \tag{2.17}$$

There exists a dimension M, such that $||R_{M+1}|| < \epsilon$ and the algorithm gives the following approximation for f

$$f = \sum_{n=1}^{M} \frac{\langle R_n, \alpha_{l_n} \rangle}{\| \phi_n \|^2} \phi_n. \tag{2.18}$$

To expand f over the original dictionary $\{\alpha_{l_n}\}_{1 \leq n \leq M}$ we must perform a change of basis. Every ϕ_n is expanded in $\{\alpha_{l_k}\}_{1 \leq k \leq M}$ as:

$$\phi_n = \sum_{k=1}^{M} b_k^{(n)} \alpha_{l_k} \quad ; \quad n = 1, \ldots, M. \tag{2.19}$$

where the coefficients $b_k^{(n)}$ are obtained by inverting the system. Inserting (2.19) in the equation (2.18), gives:

$$f = \sum_{k=1}^{M} c_k^{(M)} \alpha_{l_k}, \tag{2.20}$$

with the coefficients $c_k^{(M)}$ given as:

$$c_k^{(M)} = \sum_{n=1}^{M} b_k^{(n)} \frac{\langle R_n, \alpha_{l_n} \rangle}{\| \phi_n \|^2}. \tag{2.21}$$

The OMP approach improves the MP convergence rate and therefore amounts to a better approximation of a signal after a finite number of iterations. However, at each iteration the OMP keeps selecting the dictionary atom as prescribed by the MP approach (c.f. eq (2.9) and (2.14)), although such a selection is no longer an optimal one. In the next section we describe an approach that overcomes this limitation.

2.4 Optimised Orthogonal Matching Pursuit (OOMP)

The Optimised Orthogonal Matching Pursuit (OOMP) is a technique that improves upon the MP and OMP approaches in the following sense [16]:

At each iteration the algorithm gives an approximation to the signal, that is the orthogonal projection onto the subspace generated by the selected atoms, and minimises the norm of the corresponding residual error.

Let us recall that the approximation f_k of a signal $f \in \mathcal{H}$ is the best possible approximation in a subspace V_k, only if f_k is the orthogonal projection of f onto V_k (see appendix A). Hence, if we decompose the signal into

$$f = \sum_{n=1}^{k} c_n^{(k)} \alpha_n + \tilde{R}_k, \tag{2.22}$$

in order to minimise the norm of the residual $\tilde{R}_k$ the coefficients $c_n^{(k)}$ should satisfy:

$$\sum_{n=1}^{k} c_n^{(k)} \alpha_n = \hat{P}_{V_k} f. \tag{2.23}$$

The superscript k in $c_n^{(k)}$ indicates the dependence of these coefficients on the approximation step. The OOMP approach introduces the proposal of modifying the coefficients $c_n^{(k)}$ in subsequent iterations by means of an adaptive biorthogonalisation technique which is described below.

Out of the dictionary $\mathcal{D}$ the OOMP method chooses an arbitrary atom, say α_{l_1}, and set $V_1 = \alpha_{l_1}$. The subspace of the chosen atoms at iteration k is denoted as V_k, and at each iteration, a new subspace is constructed as: $V_{k+1} = V_k \oplus \alpha_{l_{k+1}}$. Consequently, if we denote as W_{k+1} the orthogonal complement of V_k in V_{k+1}, the orthogonal projector onto V_{k+1} can be written as: $\hat{P}_{V_{k+1}} = \hat{P}_{V_k} + \hat{P}_{W_{k+1}}$. Thus, since by definition $\alpha_{l_{k+1}} \in V_{k+1}$, the orthogonal projection of $\alpha_{l_{k+1}}$ onto W_{k+1} is the function $\psi_{k+1} = \alpha_{l_{k+1}}^{\perp}$, that satisfies:

$$\begin{aligned} \psi_{k+1} = \hat{P}_{W_{k+1}} \alpha_{l_{k+1}} &= \hat{P}_{V_{k+1}} \alpha_{l_{k+1}} - \hat{P}_{V_k} \alpha_{l_{k+1}} \\ &= \alpha_{l_{k+1}} - \hat{P}_{V_k} \alpha_{l_{k+1}}. \end{aligned} \tag{2.24}$$

The normalised to unity version of function ψ_{k+1} is denoted by $\tilde{\psi}_{k+1}$, hence we have:

$$\tilde{\psi}_{k+1} = \frac{\psi_{k+1}}{\| \psi_{k+1} \|} \tag{2.25}$$

and the representation of the corresponding orthogonal projector operator onto W_{k+1} is given by

$$\hat{P}_{W_{k+1}} f = \tilde{\psi}_{k+1} \langle \tilde{\psi}_{k+1}, f \rangle.$$

In order to obtain the coefficients $c_n^{(k+1)}$, which at iteration $k+1$ render an approximation $f_{k+1} = \hat{P}_{V_{k+1}} f$, the OOMP approach introduces a representation for the operator $\hat{P}_{V_{k+1}}$ in terms of *biorthogonal functions* [16, 14].

Adaptive biorthogonalisation technique:
By setting $\beta_{l_1}^{(1)} = \alpha_{l_1} = \tilde{\psi}_1$ the functions arising from the recursive equations:

$$\beta_{l_n}^{(k+1)} = \beta_{l_n}^{(k)} - \frac{\tilde{\psi}_{k+1}}{\| \psi_{k+1} \|} \langle \alpha_{l_{k+1}}, \beta_{l_n}^{(k)} \rangle \quad ; \quad n = 1 \ldots, k$$

$$\beta_{l_{k+1}}^{(k+1)} = \frac{\tilde{\psi}_{k+1}}{\| \psi_{k+1} \|}. \tag{2.26}$$

have been shown to fulfil the following properties [16, 14]:

a) are biorthogonal to functions α_{l_n} ; $n = 1, \ldots, k+1$, i.e,

$$\langle \beta_{l_n}^{(k+1)}, \alpha_{l_m} \rangle = \delta_{l_n, l_m} \quad ; \quad n = 1, \ldots, k+1 \quad ; \quad m = 1, \ldots, k+1$$

b) they provide a representation of the orthogonal projection as given by

$$\hat{P}_{V_{k+1}} f = \sum_{n=1}^{k+1} \langle \beta_{l_n}^{(k+1)}, f \rangle \alpha_{l_n}. \tag{2.27}$$

Due to the relevance of property b) to our purpose, the proof is given in appendix B. The biorthogonal functions (2.26) are used to recalculate the coefficients in (2.22). At each iteration the coefficients of the linear expansion representing the signal at best in the given subspace are calculated as:

$$c_{l_n}^{(k+1)} = c_{l_n}^{(k)} - \langle \beta_{l_n}^{(k)}, \alpha_{l_{k+1}} \rangle \langle \frac{\tilde{\psi}_{k+1}}{\| \psi_{k+1} \|}, f \rangle \quad ; \quad n = 1, \ldots, k$$

$$c_{l_{k+1}}^{(k+1)} = \langle \frac{\tilde{\psi}_{k+1}}{\| \psi_{k+1} \|}, f \rangle \tag{2.28}$$

with $c_1^{(1)} = \langle \alpha_{l_1}, f \rangle$.

As already stated, the OOMP differs from the MP and OMP method in the criterion to choose the atoms for the iterative approximations. Rather than selecting the atom as prescribed in (2.9) and (2.14) the OOMP aims at selecting the atom minimising the residual error $\tilde{R}_{k+1}$ in (2.22). The following theorem, the proof of which is given in appendix C, prescribes how the selection is to be made.

Theorem I [16]: The dictionary atom $\alpha_{l_{k+1}}$ that at iteration $k+1$ minimises the norm of the residue $\tilde{R}_{k+1}$ is the one yielding a maximal value of the functionals e_n ; $n = 1, \cdots, N$, where:

$$e_n = \frac{|\langle \alpha_n, \tilde{R}_k \rangle|^2}{1 - \langle \alpha_n, \hat{P}_{V_k} \alpha_n \rangle} = \frac{|\langle \psi_n, f \rangle|^2}{\| \psi_n \|^2}. \tag{2.29}$$

We sketch next the algorithm for selecting atoms by implementing condition (2.29), and successively adapting the corresponding coefficients yielding the orthogonal projection of the signal onto the selected subspace.

Set initialy $k = 1,\ \alpha'_n = \alpha_n,\ d_n = 1,\ l_1 = n\ ;\ \ n = 1, \cdots, N$ where l_1 is the index n for which $|\langle \alpha_n, f \rangle|$ has the maximal value. Set a tolerance parameter $\epsilon > 0$ and $\psi_1 = \alpha_{l_1},\ \beta_1 = \alpha_{l_1}, c_1 = \langle \alpha_{l_1}, f \rangle, \tilde{R}_0 = f,\ \ \| \tilde{R}_1 \|^2 = \| \tilde{R}_0 \|^2 - |c_1|^2$.

Step 1: For $n = 1, \ldots, N$ calculate:

$$\begin{aligned} \alpha'_n &= \alpha'_n - \psi_k \langle \psi_k, \alpha'_n \rangle \\ b_n &= \langle \alpha'_n, f \rangle \\ d_n &= d_n - |\langle \psi_n, \alpha'_n \rangle|^2 = \| \alpha'_n \|^2 \end{aligned}$$

if $|b_n| < \epsilon$ set $e_n = 0$ else $e_n = \frac{|b_n|^2}{d_n}$.

Step 2: Update: $k = k + 1,\ l_k = n$, where n is the index for which e_n is maximised.

$$\begin{aligned} \| \tilde{R}_k \|^2 &= \| \tilde{R}_{k-1} \|^2 - e_{l_k} \\ \psi_k &= \frac{\alpha'_{l_k}}{\sqrt{d_{l_k}}} \\ \beta_k &= \frac{\alpha'_{l_k}}{d_{l_k}} \\ c_k &= \langle \beta_k, f \rangle \end{aligned}$$

Step 3: For $n = 1, \ldots, k-1$ compute the biorthogonal functions, and the coefficients as:

$$\begin{aligned} \beta_n &= \beta_n - \beta_k \langle \alpha_{l_k}, \beta_n \rangle \\ c_n &= c_n - \overline{\langle \alpha_{l_k}, \beta_n \rangle} c_k, \end{aligned}$$

where with $\overline{\langle \alpha_{l_k}, \beta_n \rangle}$ we denote the complex conjugate of $\langle \alpha_{l_k}, \beta_n \rangle$.

Step 4: If $||\tilde{R}_k||^2 \leq \epsilon$ stop, else repeat steps $1 \rightarrow 4$.

Note: The above algorithm uses the Modified Gram-Schmidt technique, for constructing the orthogonal projectors $\hat{P}_{V_k}$. In some situations to avoid accumulative errors, further reorthogonalisation may be needed [20]. Alternatively we can use the QR decomposition to compute the projectors. This is actually the approach that we have used for implementing this technique in our experiments.

2.5 Redundancy Elimination Technique

Let us stress that the recursive equations (2.26) in section 2.4, can be applied on any given linear independent set of atoms. The way in which the OOMP technique selects atoms (in general, from a redundant dictionary), guarantees that such atoms are independent. There are situations, however, in which we do not have information on the signal itself, but only on the subspace it may lie in. In some cases, it is convenient to have a representation of the subspace by means of the minimum possible numbers of atoms. To this end we need to eliminate redundancy, in other words: we need to build bases for the subspace.

Here we will adopt the technique proposed in [15] to construct dictionaries of linearly independent atoms and the corresponding biorthogonal functions.

Note that from the definition of the functions ψ_k (cf. (2.24)) it follows that linearly dependent atoms give rise to functions ψ_k of zero norm. Hence, simply by disregarding those atoms, we can select an independent set. Nevertheless, in practice we have to deal with dictionaries yielding several functions ψ_n of small norm, thereby producing important numerical errors. In order to reduce numerical errors, it is proposed in [15] that the selection should be made in the following hierarchical way: at iteration k the atom α_n maximising the norm of the corresponding function $\psi_n = \alpha_n - \hat{P}_{V_{k-1}}\alpha_n$ should be selected, i.e. at iteration k the selected atom α_{l_k} is the one yielding a maximum value of the following quantities:

$$||\psi_n||^2 = 1 - \langle \alpha_n, \hat{P}_{V_{k-1}}\alpha_n \rangle.$$

Given a tolerance parameter $\epsilon > 0$ let us assume that the situation for which all values of $||\psi_n||^2$ are less than ϵ, is reached at iteration k. Hence, the method has selected the k linearly independent atoms, up to the given tolerance. In addition to the linear independent atoms α_{l_j} ; $j = 1, \ldots, k$ the algorithm provides us with the corresponding biorthogonal functions β_{l_j} ; $j = 1, \ldots, k$, computed as prescribed in (2.26).

Let us assume now that we are given a signal $f \in \mathcal{H}$ and we choose the selected basis α_{l_j} ; $j = 1, \ldots, k$ to represent it. Since the biorthogonal functions yielding the representation of the orthogonal projector onto V_k are already computed, the atomic decomposition of the signal

$$f_k = \sum_{n=1}^{k} c_n^{(k)} \alpha_{l_n}, \tag{2.30}$$

is readily obtained from the coefficients computed as the inner products:

$$c_n^{(k)} = \langle \beta_n^{(k)}, f \rangle \quad ; \quad n = 1, \ldots, k. \tag{2.31}$$

2.6 The Backward Optimised Orthogonal Matching Pursuit (BOOMP)

The biorthogonalisation technique we have discussed in section 2.4 allows for the modification of biorthogonal functions in order to account for an additional atom in the spanning set. As already discussed, the coefficients of the atomic decomposition (2.22) need also to be recalculated if one atom is removed from the set. This can be achieved by a backward biorthogonalisation technique proposed in [12]. Based on this technique an approach for reducing coefficients in an atomic decomposition is reported in [13]. Such an approach is termed Backward Optimised Orthogonal Matching Pursuit (BOOMP), because it selects the atoms to be disregarded according to a criterion which is equivalent to the one proposed by the OOMP technique.

Given an atomic decomposition of a signal $f \in \mathcal{H}$ through the linear expansion:

$$f = \sum_{n=1}^{k} c_n^{(k)} \alpha_n, \tag{2.32}$$

where the atoms α_n are assumed to be linearly independent, the BOOMP approach provides a set of recursive equations for modifying the coefficients $c_n^{(k)}$ when one of these coefficients is to be disregarded. The approach decides on the atom to be disregarded by minimising the norm of the residual error.

Let us assume that we are given a dictionary of k linearly independent atoms and let $V_k = \text{span}\{\alpha_n; n = 1, \ldots, k\}$. Let us further assume that the corresponding functions $\{\beta_n^{(k)} \; ; \; n = 1, \ldots, k\}$ representing $\hat{P}_{V_k}$ are known. Hence the approximation in V_k of an arbitrary signal $f \in \mathcal{H}$ is obtained as

$$\hat{P}_{V_k} f = \sum_{n=1}^{k} \alpha_n \langle \beta_n^{(k)}, f \rangle = \sum_{n=1}^{k} c_n^{(k)} \alpha_n. \tag{2.33}$$

with the coefficients given by the the inner products:

$$c_n^{(k)} = \langle \beta_n^{(k)}, f \rangle. \tag{2.34}$$

Consider now that the atom α_j is to be removed from V_k and denote the subspace of the remaining atoms as $V_{k/\alpha_j} = \text{span}\{\alpha_1, \ldots, \alpha_{j-1}, \alpha_{j+1}, \ldots, \alpha_k\}$. Then, the optimal approximation of the signal will be given by:

$$\hat{P}_{V_{k/\alpha_j}} f = \sum_{\substack{n=1 \\ n \neq j}}^{k} \alpha_n \langle \beta_n^{(k/j)}, f \rangle = \sum_{\substack{n=1 \\ n \neq j}}^{k} c_n^{(k/j)} \alpha_n, \tag{2.35}$$

It is proved in [12] that the functions $\beta_n^{(k/j)}$ in (2.35) can be obtained from the recursive formula:

$$\beta_n^{(k/j)} = \beta_n^{(k)} - \frac{\beta_j^{(k)} \langle \beta_j^{(k)}, \beta_n^{(k)} \rangle}{\| \beta_j^{(k)} \|^2} \quad ; \quad n = 1, \cdots, j-1, j+1, \cdots, k. \tag{2.36}$$

Consequently, the coefficients $c_n^{(k/j)}$ in (2.35) are computed as

$$c_n^{(k/j)} = c_n^{(k)} - \frac{\langle \beta_n^{(k)}, \beta_j^{(k)} \rangle}{\| \beta_j^{(k)} \|^2} c_j^{(k)}. \tag{2.37}$$

Now that we know how to modify the remaining coefficients when one coefficient, say c_j, is to be omitted, we need to consider a criterion for choosing c_j. The following theorem establishes the BOOMP selection criterion. The proof can be found in appendix D.

Theorem II: Let $\tilde{R}_j$ be the residual resulting by disregarding the coefficient $c_j^{(k)}$ as passing from approximation f_{V_k} to $f_{V_{k/\alpha_j}}$ i.e., $f_{V_k} = f_{V_{k/\alpha_j}} + \tilde{R}_j$. In order to minimise the norm of the residual $\tilde{R}_j$ such coefficient is to be chosen as the one yielding a minimum value of the quantity

$$\frac{|c_j^{(k)}|^2}{||\beta_j^{(k)}||^2}. \tag{2.38}$$

Assuming that the sets $\{\alpha_n \ ; \ n = 1, \ \ldots \ , k\}$ and $\{\beta_n^{(k)} \ ; \ n = 1, \ \ldots \ , k\}$ are known, the BOOMP disregards coefficients of an atomic decomposition as prescribed by the following steps:

Step 1: Compute the coefficients $c_n^{(k)} = \langle \beta_n^{(k)}, f \rangle$, and find the index j for which the expression $\frac{|c_j^{(k)}|^2}{||\beta_j^{(k)}||^2}$ is minimised.

Step 2: Modify the biorthogonal functions using equation (2.36), and the remaining coefficients using (2.37).

Step 3: Repeat steps 1 and 2 until a given tolerance error for the approximation is reached.

2.7 Numerical example on MP algorithms

2.7.1 Aims

We illustrate here, numerically, some typical features concerning the performance of the MP, OMP and OOMP algorithms introduced in the sections 2.2, 2.3 and 2.4. The posterior application of BOOPM leading to interesting conclusions.

2.7.2 Procedure

We will approximate the signal of figure 2.1 (left graph), using a dictionary of mexican hat wavelets, which will be introduced in section 4.2.

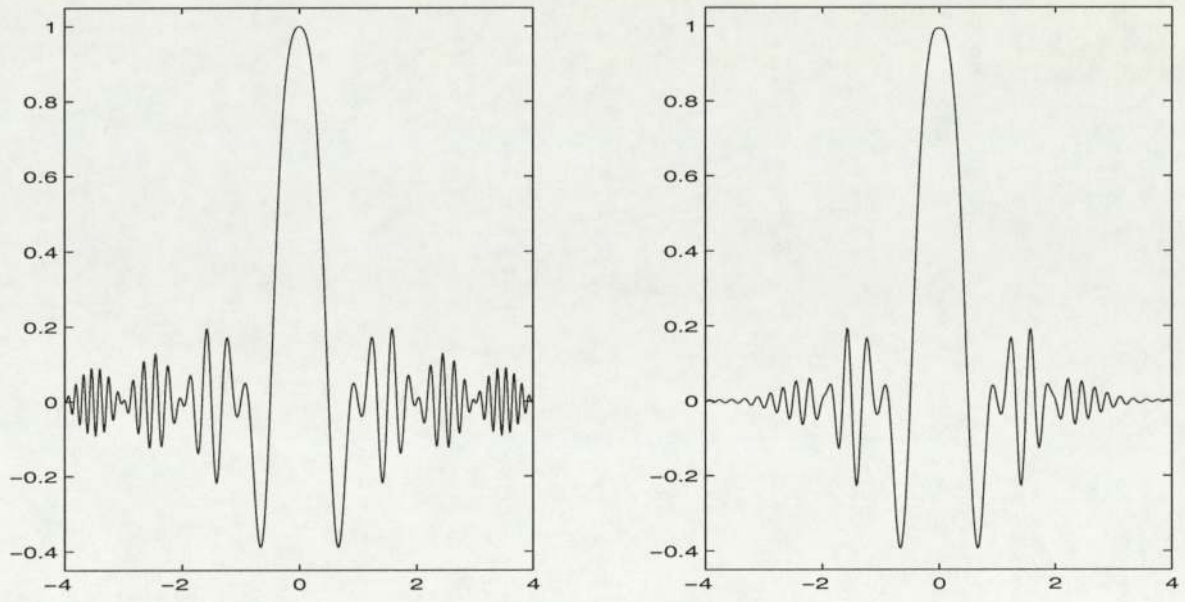

Figure 2.1: Left Graph: The chirp signal. Right Graph: Approximation of the chirp with mexican hat wavelets, using the MP approach.

First we apply the MP approach. This approach exhibits slow convergence rate, after 10000 iterations the representation of the signal is the one of figure 2.1 (right graph). A graph of the error for the first 3000 iterations, is given in figure 2.2.

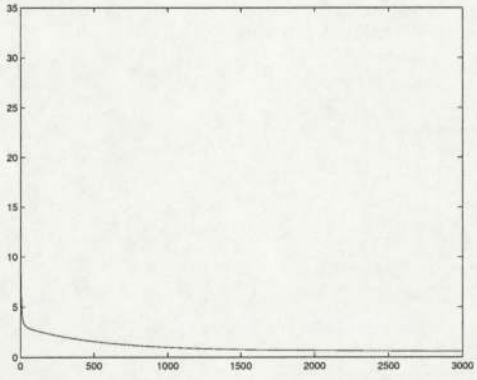

Figure 2.2: Approximation error of the chirp with mexican hat wavelet, using MP (10000 iterations).

We also apply the OMP within the workings of the OOMP approach, i.e. computing biorthogonal functions rather than inverting a matrix, as originally proposed in [11, 6]. For this example we have found that the OMP needs 25 more coefficients than the OOMP approach to produce an approximation of the same quality. In figure 2.3 (top

left graph) it is plotted the OMP approximation with 100 coefficients. The OOMP approximation for the same number of coefficients is plotted in figure 2.3 (top right graph). See table 2.1 for the corresponding values of residual errors.

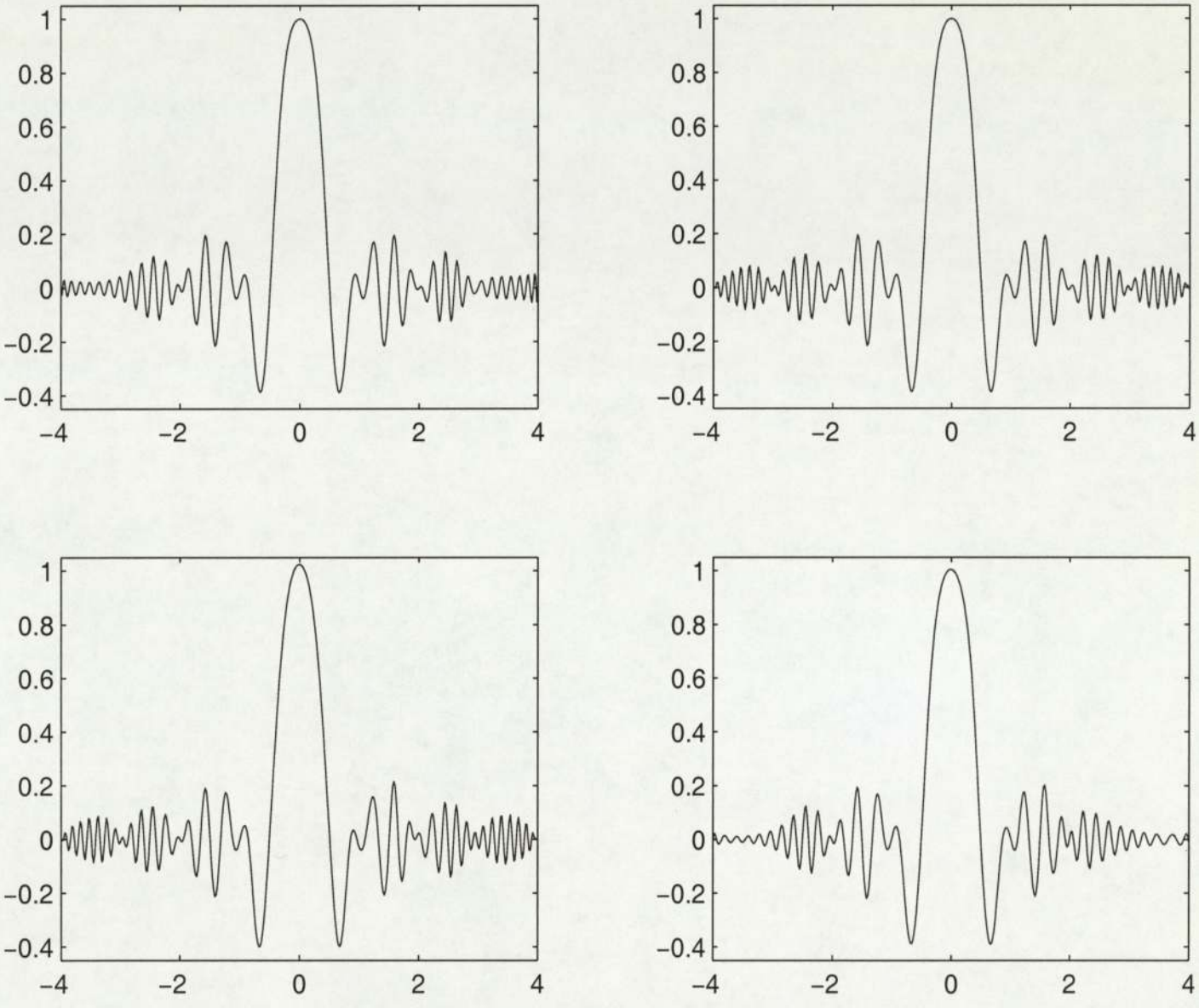

Figure 2.3: Top Left Graph: Approximation of the chirp signal with the mexican hat dictionary using the OMP approach and 100 coefficients. Top Right Graph: Approximation using the OOMP and 100 coefficients. Bottom Left Graph: Applying the BOOMP to the previous approximation, 75 coefficients. Bottom Right Graph: Approximation using OOMP, 75 coefficients

We now apply the BOOMP to the OOMP approximation of figure 2.3. We see that disregarding 25 coefficients the approximation is still acceptable (bottom left graph, figure 2.3). Nevertheless if the OOMP is stopped after 75 iterations the resulting approximation is the one plotted in the bottom right graph, of the same figure.

Coefficients	OMP	OOMP
100	9.64828e-4	2.2229e-04
125	2.2165e-4	2.7681e-05

Table 2.1: Error of the approximation using the OOMP and the MP algorithms.

2.8 Remarks

In this chapter we have discussed some techniques for adaptive approximation of a signal in terms of non-orthogonal atoms. The differences of the MP, OMP and OOMP approaches have been established from a formal viewpoint. In addition, these approaches have been compared in relation to convergence rate by recourse to a numerical example. The results are in accordance with what can be expected from the theoretical considerations: OOMP renders the fastest convergence rate. In the given example, the difference as compared with the MP approach being enormous. The difference with OMP being less pronounced, however, since the implementation of OMP is not less computational demanding there is no practical advantage of this approach over OOMP. Hence, in all our relevant experiments we will use only the OOMP approach. A really interesting result arises from the application of BOOMP to the OOMP approximation. It strongly suggests that a smart forward/backward procedure could improve sparseness in the representation. Unfortunately, due to time limitation we have not been able to implement such a technique. We leave this line of research as one of our proposals of future work.

Chapter 3

The Haar based dictionary

3.1 Introduction

In this chapter we introduce a non-orthogonal dictionary, which is generated from the well known orthogonal Haar wavelet system. We discuss the Haar basis in [0,1] and transform this basis into a redundant set. We show that, in the finite dimensional case, the new set spans a subspace larger than the one expanded by the orthogonal wavelets. Experiments illustrate potential advantages of the proposed non-orthogonal system.

3.2 Dyadic intervals

Definition: For each pair of integers $m, n \in \mathbb{Z}$ we define the interval $I_{m,n}$ by:

$$I_{m,n} = [2^{-m}n, 2^{-m}(n+1)).$$

The collection of all such intervals is called the collection of *dyadic intervals* on R.

One useful property of the dyadic intervals, is the following:

Lemma: Given $m_0, n_0, m_1, n_1 \in \mathbb{Z}$ and either $m_0 \neq m_1$ or $n_0 \neq n_1$, then one of the following must be true:

- $I_{m_1,n_1} \cap I_{m_0,n_0} = \emptyset$
- $I_{m_1,n_1} \subseteq I_{m_0,n_0}$
- $I_{m_0,n_0} \subseteq I_{m_1,n_1}$.

In the last two cases, the smaller interval is contained in either the left or the right half of the larger interval.

Another notation we can use, based on the previous lemma, is the following:

Definition: Given a dyadic interval of scale m, $I_{m,n}$ we write:

$$I_{m,n} = I^l_{m,n} \cup I^r_{m,n},$$

where $I^l_{m,n}$ is the left and $I^r_{m,n}$ the right half of the interval $I_{m,n}$ of scale $m+1$.

3.3 The Haar system on R

Let us define the function $X_{[a,b)}$ as:

$$X_{[a,b)}(x) = \begin{cases} 1, & \text{if} \quad x \in [a,b) \\ 0, & \text{elsewhere} \end{cases}$$

Let $p(x) = X_{[0,1)}(x)$, and for each $m, n \in \mathbb{Z}$, we define

$$p_{m,n}(x) = 2^{\frac{m}{2}} p(2^m x - n).$$

The collection $\{p_{m,n}(x)\}_{n\in\mathbb{Z}}$ is referred to as the system of scale m *Haar scaling functions.*

Let $h(x) = X_{[0,1/2)}(x) - X_{[1/2,1)}(x)$ (see figure 3.1), and for each $m, n \in \mathbb{Z}$ define

$$h_{m,n}(x) = 2^{\frac{m}{2}} h(2^m x - n).$$

The collection $\{h_{m,n}(x)\}_{m,n\in\mathbb{Z}}$ is referred to as the *Haar system on R.* For each $m \in \mathbb{Z}$ the collection $\{h_{m,n}(x)\}_{n\in\mathbb{Z}}\}$ is referred to as the system of scale m Haar functions. We call m the *scale parameter* and n the *translation parameter.* The Haar system is created through translations and dilations of the function $h(x)$, which we call *mother wavelet.*

Some important remarks on the Haar system on R:

- For each $m, n \in \mathbb{Z}$

  $$h_{m,n}(x) = 2^{\frac{m}{2}}(X_{I^l_{m,n}} - X_{I^r_{m,n}}).$$

  So the $h_{m,n}$ is supported on the interval $I_{m,n}$ and does not vanish on that interval. We say in this case, that the Haar function $h_{m,n}$ is *associated* with the interval $I_{m,n}$.

- For each $m, n \in \mathbb{Z}$ we have:

  $$\int_R h_{m,n}(x)dx = \int_{I_{m,n}} h_{m,n}(x)dx = 0$$

  $$\int_R |h_{m,n}(x)|^2 dx = \int_{I_{m,n}} |h_{m,n}(x)|^2 dx = 1.$$

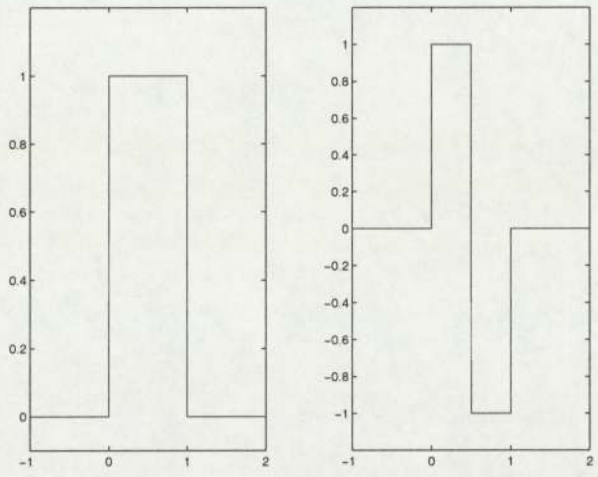

Figure 3.1: The Haar scaling function $p_{0,0}$(left graph) and the mother wavelet for the Haar system $h_{0,0}$(right graph).

From the above remarks and using the properties of the dyadic intervals, we conclude that:

The Haar system on R is an orthonormal *system on R.*

We now define the Haar system on [0,1], which is of relevance to our experiments.

3.4 The Haar system on [0,1]

Definition: For any integer $J \geq 0$ the scale J Haar system on [0,1] is the collection:

$$\{p_{J,n}(x) : 0 \leq n \leq 2^J - 1\} \cup \{h_{m,n}(x) : m \geq J; 0 \leq n \leq 2^m - 1\}.$$

For $J = 0$, this collection is referred to as the Haar system on [0,1].

As defined above the Haar system on [0,1] consists of precisely those Haar functions $h_{m,n}$ corresponding to *dyadic* intervals $I_{m,n}$ that are subsets of [0,1] (see figure 3.2), together with the scaling function $p_{0,0}(x)$.

In figure 3.3 we plot some elements of the Haar system in [0,1], corresponding to scales $m = 0, 1, 2$ and 3.

The following theorem is essential to our purpose. The proof is given in [19].

Theorem: For each integer $J \geq 0$, the scale J Haar system on [0,1] is a *complete orthonormal system* on [0,1].

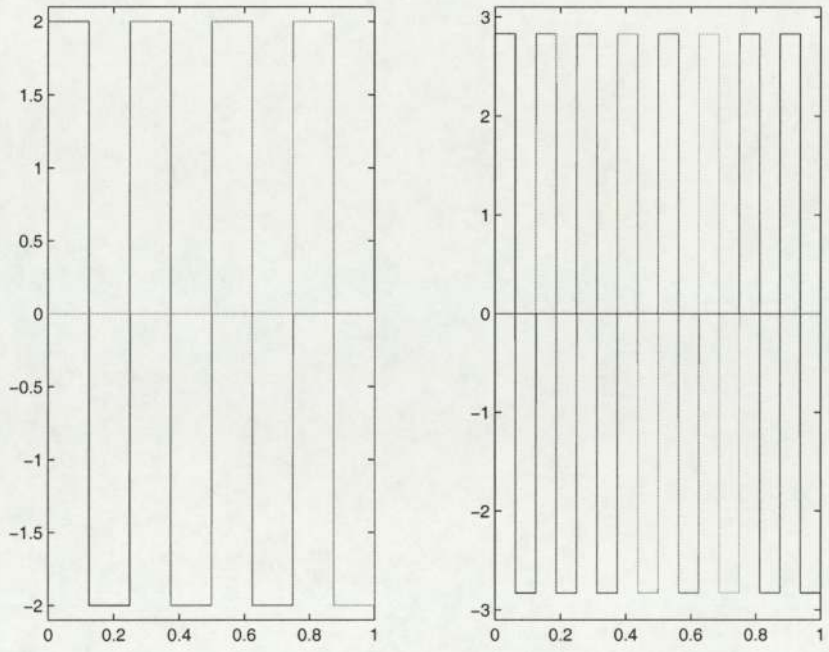

Figure 3.2: Haar wavelets on [0,1] corresponding to scales $m = 2$ (left graph) and $m = 3$ (right graph).

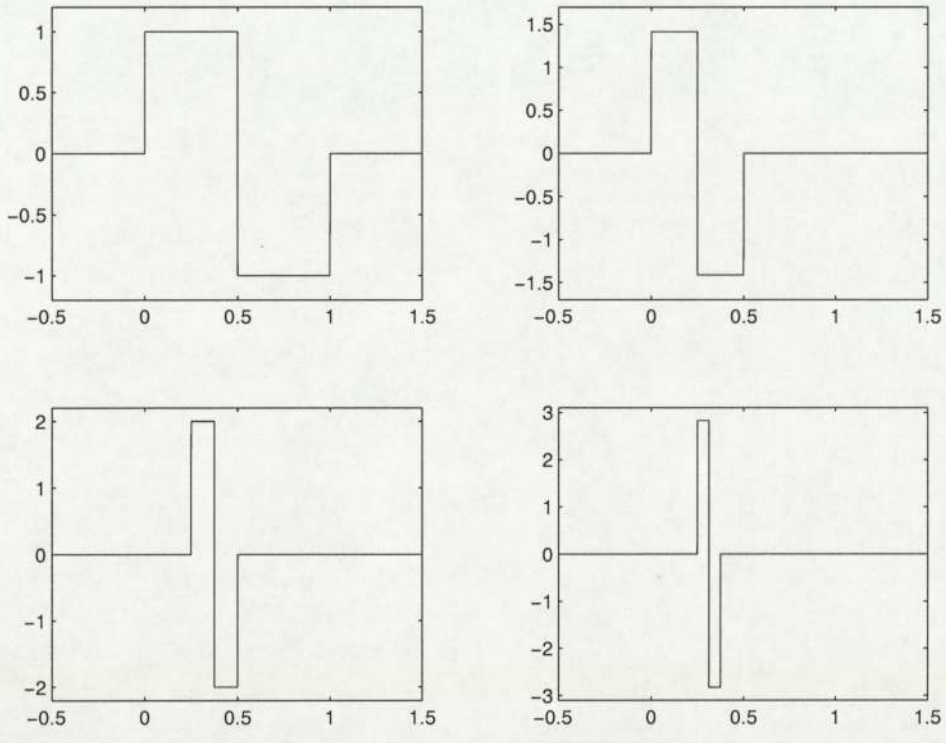

Figure 3.3: The Haar wavelets for scales $m = 0$ to 3 from top left to bottom right.

The subspace spanned by the single function $p_{0,0}(x)$ will be called V_0 and the $span\{h_{m,n}; 0 \leq n \leq 2^m - 1\}$ will be called W_m. We also define the subspace

$$V_m = V_0 \cup W_1 \cup \cdots \cup W_m. \tag{3.1}$$

3.5 Non-orthogonal Haar based dictionary

In order to introduce our non-orthogonal Haar based system let us recall that with $h(x)$ we denote the mother Haar wavelet function and the set of functions $\{p_{m,n}(x)\}_{n\in\mathbb{Z}}$, is the scale m Haar scaling functions system. We introduce now the key ingredient for our construction: rather than considering the translation parameter an integer (as required by the orthogonality condition) we let this parameter be of the form $b \cdot n$ with $n \in \mathbb{Z}$ and b a real number in $(0, 1)$. Hence the corresponding wavelet functions are of the form:

$$\alpha_{m,n}(x) = 2^{\frac{m}{2}} h(2^m x - bn). \tag{3.2}$$

The following definition completely characterises our Haar based dictionary on [0,1].

Definition: The non-orthogonal Haar based system on [0,1] is the collection:

$$p_{0,0}(x) \cup \{\alpha_{m,n}(x) : \ m \geq 0, \ 0 \leq n \leq 2^m - \frac{1}{b}\}.$$

In figure 3.4 some wavelets of the non-orthogonal system are plotted, for scale $m = 1$ and different values of the translation parameter.

We will denote the scale m non-orthogonal Haar based system as $R_m = \text{span}\{\alpha_{m,n}; 0 \leq n \leq 2^m - \frac{1}{b}\}$ and the non-orthogonal Haar based system from scale 0 to scale j will be denoted as:

$$B_j = V_0 \cup R_1 \cup \cdots \cup R_j. \tag{3.3}$$

As a consequence of introducing the parameter b in the wavelets definition (cf. (3.2)), the non-orthogonal dictionary contains atoms $\alpha_{m,n}$ that are no longer supported on the dyadic interval $I_{m,n}$, i.e., we can have wavelets that vanish on a given dyadic interval. It is clear then that, by relaxing the orthogonality condition we have, for each scale, a larger number of atoms, which completely lie inside the [0,1] interval. This is illustrated in figure 3.5.

Since the orthogonal Haar system on [0,1] is a complete set of functions, it is clear that by incorporating more functions in the whole set we can only introduce redundancy. The question arises, however, as to whether this is also true when considering

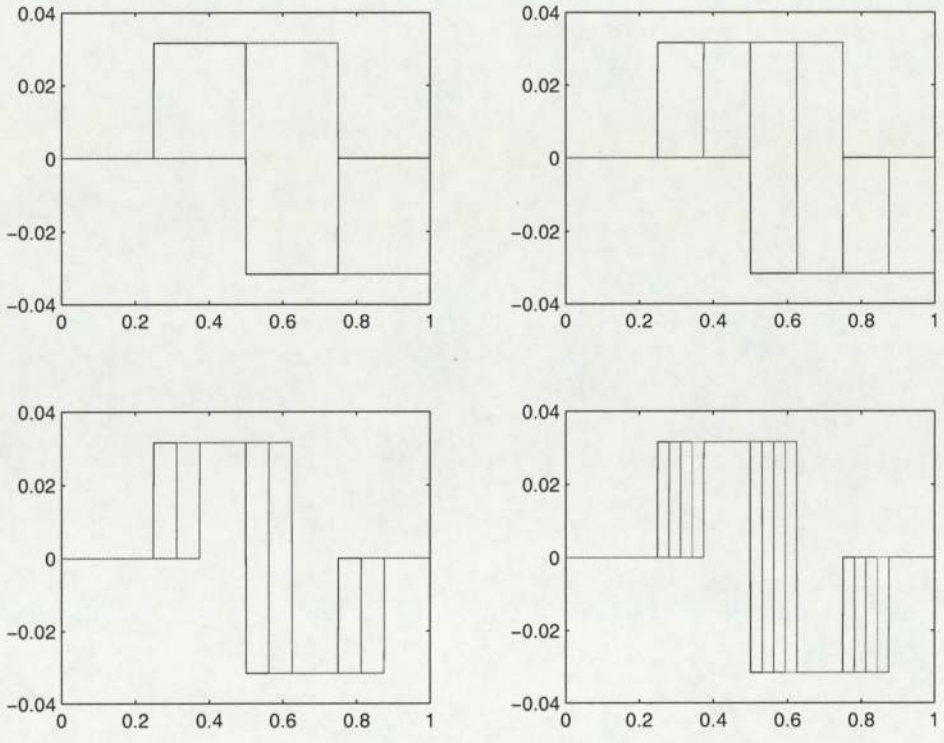

Figure 3.4: Non-orthogonal Haar wavelets for scale $m = 1$ and for translation parameters $b = 0.5,\ 0.25,\ 0.125$ and 0.0625 (from top left to bottom right)

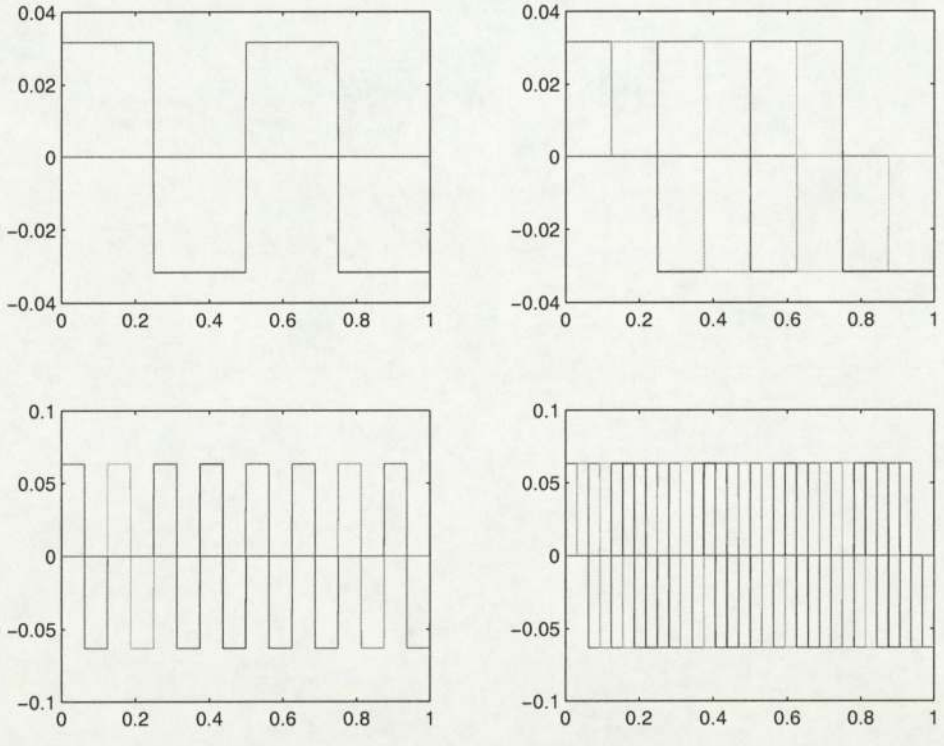

Figure 3.5: Orthogonal Haar wavelets corresponding to scales $m = 1$ and $m = 2$ (left column), and non-orthogonal Haar based wavelets for the same scales and $b = 0.25$ (right column).

a finite subset of functions, up to scale m say. In the next section we present an experiment that is specially devised to answer this question.

3.6 Thought experiment

3.6.1 Aims

The aim of this experiment is to show that if we consider the linear span of Haar functions up to a fixed scale, the incorporation of non-orthogonal Haar functions, defined in section 3.5, up to the same scale, not only incorporates redundancy, but extends the subspace.

We consider the subspaces V_j and B_j as defined in (3.1) and (3.3) respectively. Since V_j is the subspace spanned by the Haar functions from scale 0 to j, and B_j is the subspace spanned by the non-orthogonal Haar functions within the same scales, by construction we have that V_j is included in B_j. In order to show that $V_j \neq B_j$ it is enough to show that there is a signal in B_j which does not belong to V_j.

3.6.2 Procedure

We consider for this experiment the subsets V_3 and B_3. We define a signal $f \in \mathcal{H}$ in the interval [0,1] as follows: Out of the dictionary B_3 we take three dictionary functions at scale $m = 3$ and $b = 0.25$. The functions are chosen in such a way that none of them is an element of V_3. We construct f as a linear combination of such functions. Hence $f \in B_3$. This signal is plotted in figure 3.6 with the solid line. If we try to approximate the signal using the orthogonal subdictionary V_3, the approximation we obtain is shown in figure 3.6 on the left. It is clear from the graph, that $f \notin V_3$.

In addition to the previous experiment we consider the dictionary V_4, which is obtained by adding to V_3 the orthogonal Haar functions of one higher scale. The representation that we obtain now coincides with the signal, i.e $f \in V_4$ (see figure 3.7).

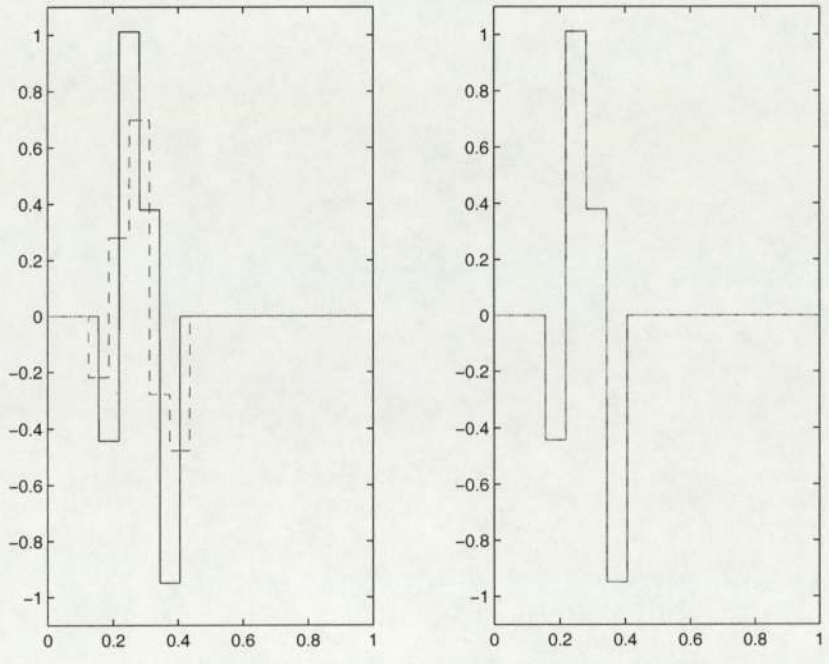

Figure 3.6: Approximation of a signal f with two subdictionaries involving the same scales. The graph on the left corresponds to the orthogonal Haar dictionary and the one on the right to the non-orthogonal one.

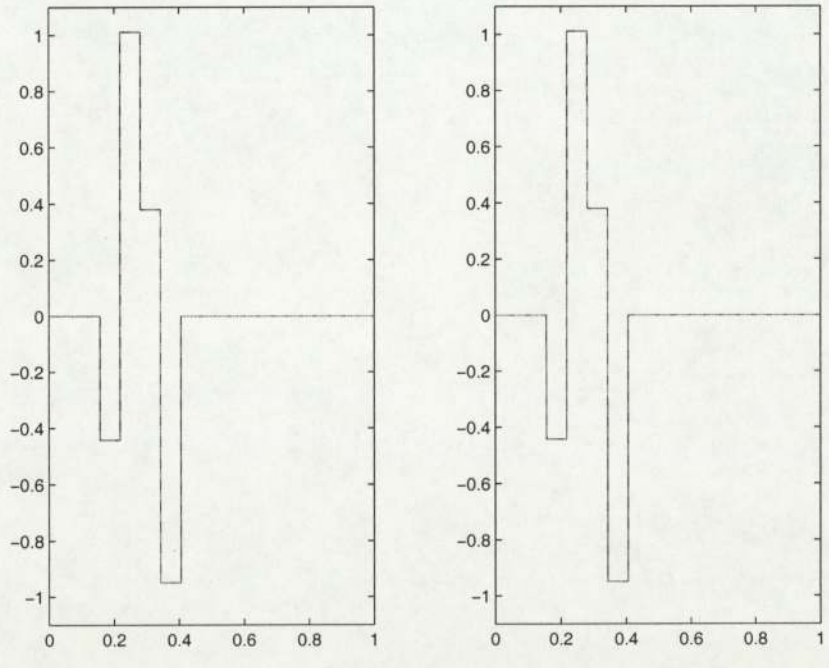

Figure 3.7: Approximation of a signal f with orthogonal Haar wavelets up to scale $m = 4$ (left graph), and approximation with the non-orthogonal Haar wavelets up to a coarser scale $m = 3$ and $b = 0.25$(right graph).

3.6.3 Conclusions

In the experiment above we have found a signal $f \in B_3$ which is not in V_3. This leads to conclude that in the finite dimensional case, by increasing the number of functions at a fixed scale we can extend the subspace generated by orthogonal functions at the same scale.

We are now in a position to answer the question motivating the experiment:

In the finite dimension case, by decreasing the translation parameter b of the non-orthogonal Haar functions we do not only incorporate redundancy. The new, non-orthogonal dictionary creates a larger subspace, that is similar to a subspace created by orthogonal dictionaries of higher scales.

From the experiments we have also seen that with the dictionary B_3 and setting the parameter $b = 0.25$ we have the same approximation as with V_4. This suggests that we can generate Haar wavelets of higher scales, by using the non-orthogonal Haar dictionaries of smaller scales, only by introducing a translation parameter $b < 1$. In the following experiments we will try to show that these conclusions hold in other cases as well, where we approximate arbitrary signals, and try to improve by changing the value of the translation parameter b.

3.7 Experiments using the non-orthogonal Haar dictionary

3.7.1 Aims

In the thought experiment given in the previous section, we concluded that by decreasing the translation parameter b of the non-orthogonal Haar functions (c.f. eq (3.2)), we introduce a larger subspace. Here we illustrate this fact further, by approximating two signals of different nature. All approximations are obtained by using the OOMP approach discussed in section 2.4.

We already know that using Haar wavelets of higher scales allows us to improve approximations. Theoretically, by increasing the scale to infinity we can represent any signal in $\mathcal{H}$. In practice, however, we cannot use arbitrary large scales, as increasing the scale implies having to increase the resolution in representing the functions. On the other hand, we have already seen that by relaxing the condition of orthogonality and considering a translation parameter $b = 0.25$ it is possible to obtain, with coarser scale, the same approximation as with an orthogonal subdictionary at a finer scale.

The aim of this series of experiments is to show that we can keep simulating finer and finer scales, by letting the translation parameter b decrease by powers of 2.

3.7.2 Procedure

For the first series of experiments we will consider the *blocky* signal of figure 3.8. We will compare the dictionary V_3 with the dictionary B_3 for different values of b. In figure 3.8 (top graph on the left), we can see the approximation we obtain by using the orthogonal dictionary V_3. The approximations obtained with the non-orthogonal dictionary are plotted in the remaining graphs of same figure, for the values $b = 0.25$, $b = 0.125$, and $b = 0.0625$. The orthogonal dictionary V_3 cannot give a better approximation for this scale. However, as shown in figure 3.8, by letting the value of b decrease by powers of 2 we have obtained better approximations without changing the scale.

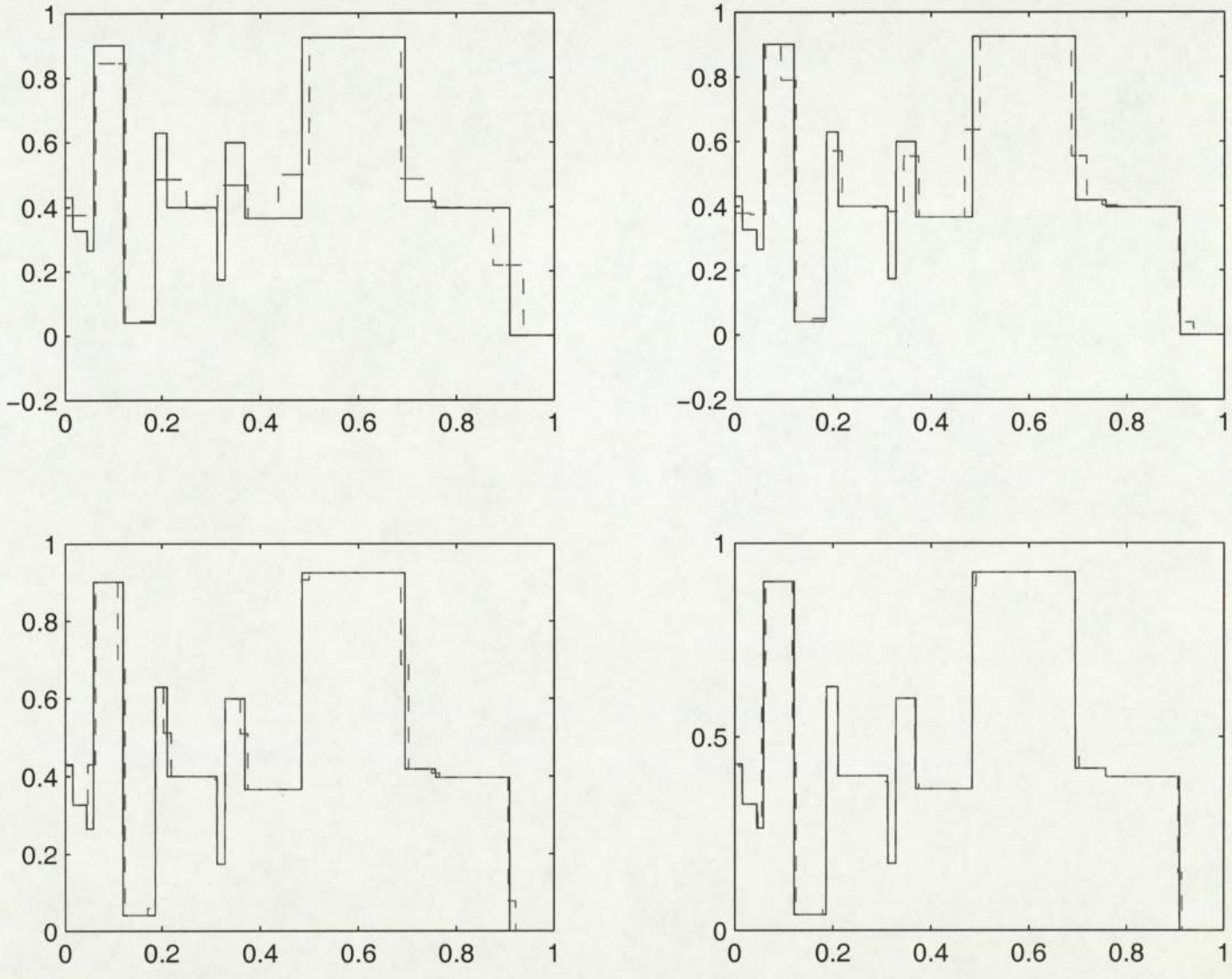

Figure 3.8: Approximation of the blocky function using the orthogonal Haar dictionary of scale $m = 3$ (top graph on the left) and with the non-orthogonal dictionaries of scale $m = 3$ and values of b equal to 0.25, 0.125, and 0.0625 respectively.

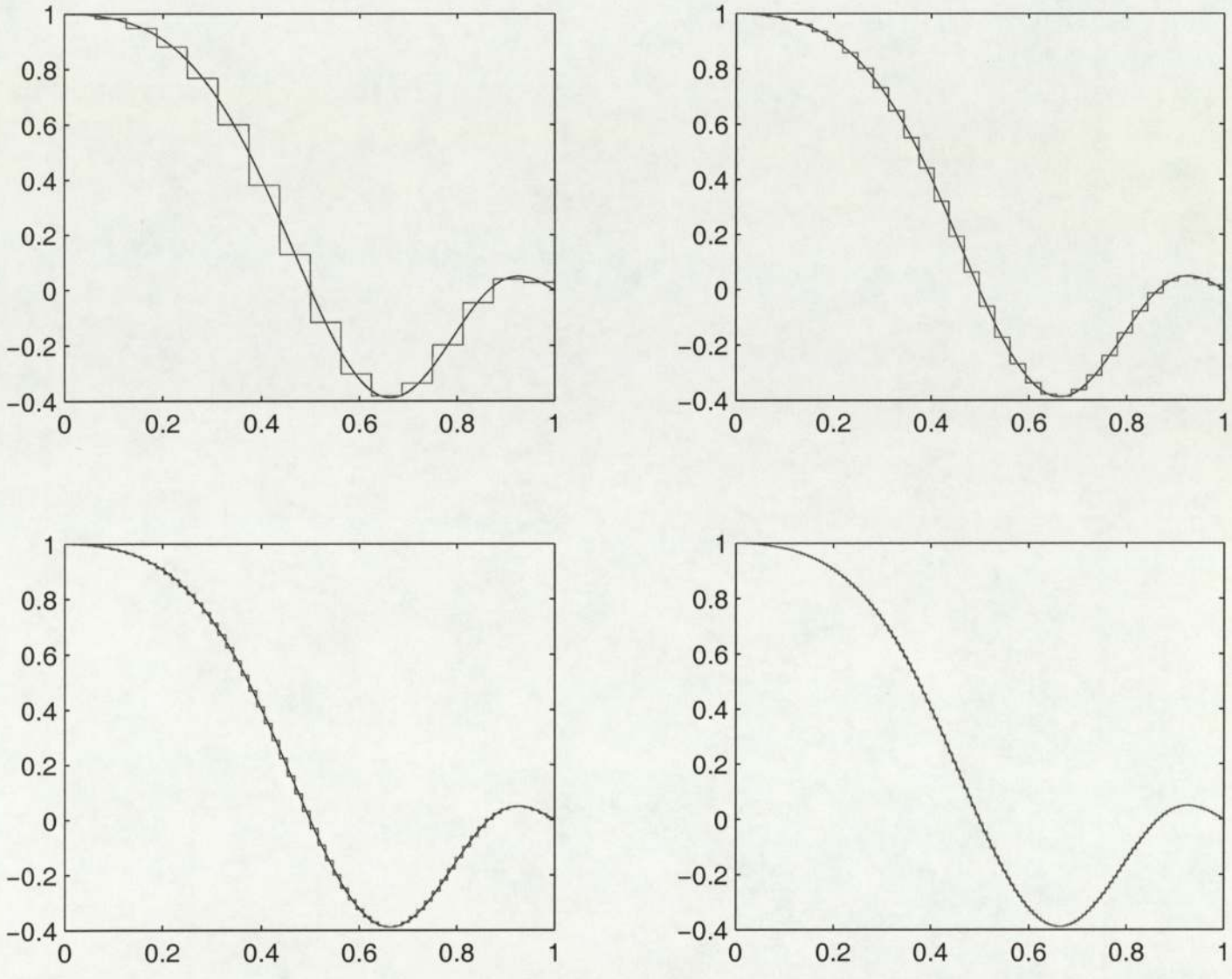

Figure 3.9: Approximation of a smooth signal by the orthogonal Haar dictionary of scale $m = 3$ (top graph on the left). Subsequent approximations by considering the non-orthogonal Haar dictionaries of the same scale and values of b equal to 0.25, 0.125, and 0.0625 respectively.

We consider now the smooth signal in figure 3.9. With an orthogonal Haar dictionary we can have a good approximation of this type of signal only by using a considerably large scale. If we consider the subdictionary V_3 the approximation is the really poor one shown in figure 3.9 (top graph on the left). As in the previous example, when using the non-orthogonal dictionary B_3, and by decreasing the value of the parameter b, we can achieve a higher quality approximation without changing the scale.

3.7.3 Conclusions

We have approximated two signals of different nature. A blocky and a smooth signal. We have used the orthogonal Haar dictionary up to scale 3, V_3, and the non-orthogonal Haar based dictionary B_3 of the same scale, but for different values of b. We have obtained good approximations of both signals, without increasing the scale, but by letting the parameter b decrease. This can only be done by resigning orthogonality. Nevertheless, the release of such a property has allowed us to approximate functions, like the one in figure 3.9, which require high scales otherwise, without having to increase the precision of the representation. From the experiments of this section we conclude that:

By decreasing the translation parameter b, we can achieve better and better approximations without changing the scale

3.8 Redundancy elimination using the non-orthogonal Haar dictionary

3.8.1 Aims

So far we have considered the problem of selecting atoms from a non-orthogonal dictionary as a signal dependent problem: Given a signal, we have iteratively selected the atoms that, at each iteration, yield an optimal approximation of the signal. Here the selection will be carried out through a signal independent procedure: Out of a non-orthogonal dictionary, we will eliminate redundancy by the method described in section 2.5. For the non-orthogonal Haar based dictionaries we are considering in this chapter, all different sets of linear independent atoms are guaranteed to span the same subspace. This is due to the fact that, in these types of of dictionaries, the number of linear independent atoms can be determined without ambiguity. Hence, by eliminating redundancy we can easily obtain different bases for the identical subspace.

The aim of the experiments in this section is to extract different bases from a non-orthogonal Haar based dictionary. The corresponding biorthogonal bases will also be computed.

Once the biorthogonal bases are available, given a signal one can immediately compute the coefficients of its representation in different bases. Moreover, by disregarding coefficients through the BOOMP approach (section 2.6), one can assess if a basis is more adequate than others for representing that given signal. The suitability of a basis may follow, in some cases, simply from the cardinality of the non-zero coefficients which are needed to represent the signal in such a basis. In other cases, entropy and other concave measures may be needed to decide on the suitability of a basis [7].

3.8.2 Procedure

In these experiments we consider the dictionary B_3 with translation parameter $b = 0.0625$. By eliminating redundancy we obtain different bases for the subspace. These bases however are not orthogonal, so that we will have to use the corresponding biorthogonal functions to obtain the representation of the signal. In figure 3.10 we can see that for an identical atom participating in two different bases, the corresponding biorthogonal function is different.

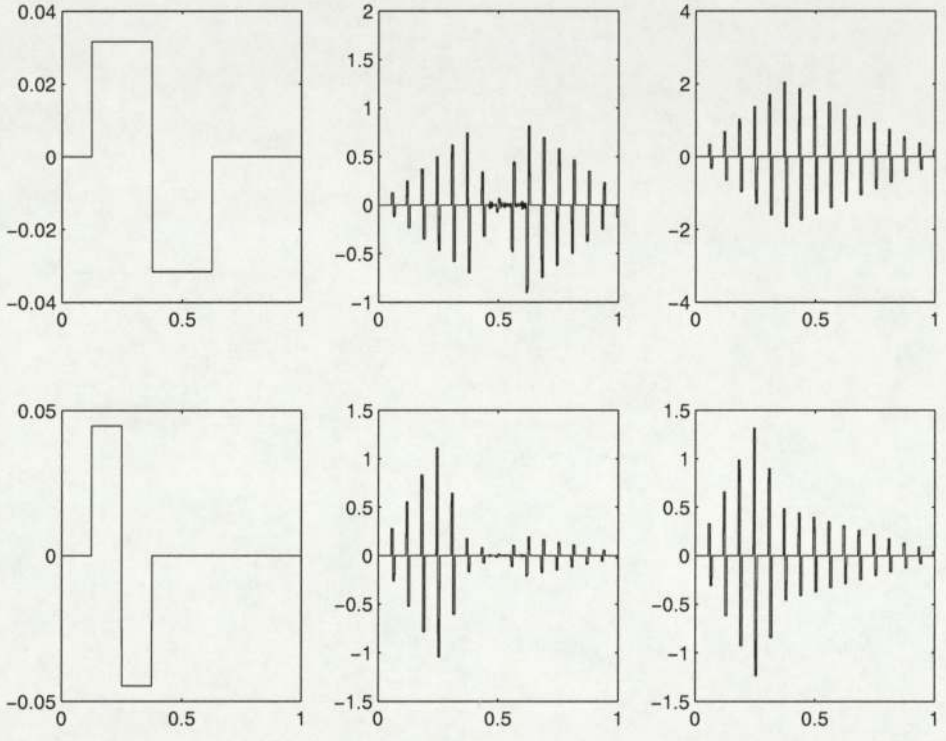

Figure 3.10: Each row corresponds to an atom and the corresponding biorthogonal functions for a different basis. Every time we choose a new basis, the biorthogonal functions are different.

Having different bases we can approximate any given signal with a subset of the bases. Using the corresponding biorthogonal basis we compute the coefficients of the atomic decomposition and then apply the BOOMP for disregarding some of them. In figure 3.11 we approximate the blocky signal using two different bases. If we use all the coefficients the approximations we obtain are exactly the same, since both are approximations in the same subspace. As shown in figure 3.12 the coefficients of each representation are, of course, different. Also after applying the BOOMP for reducing coefficients in each case, the remaining coefficients vary quite a lot (see figure 3.12, lower row).

The fact that a basis may be more appropriate than other is made clear in the following experiment. When approximating the signal of figure 3.13, we have the same

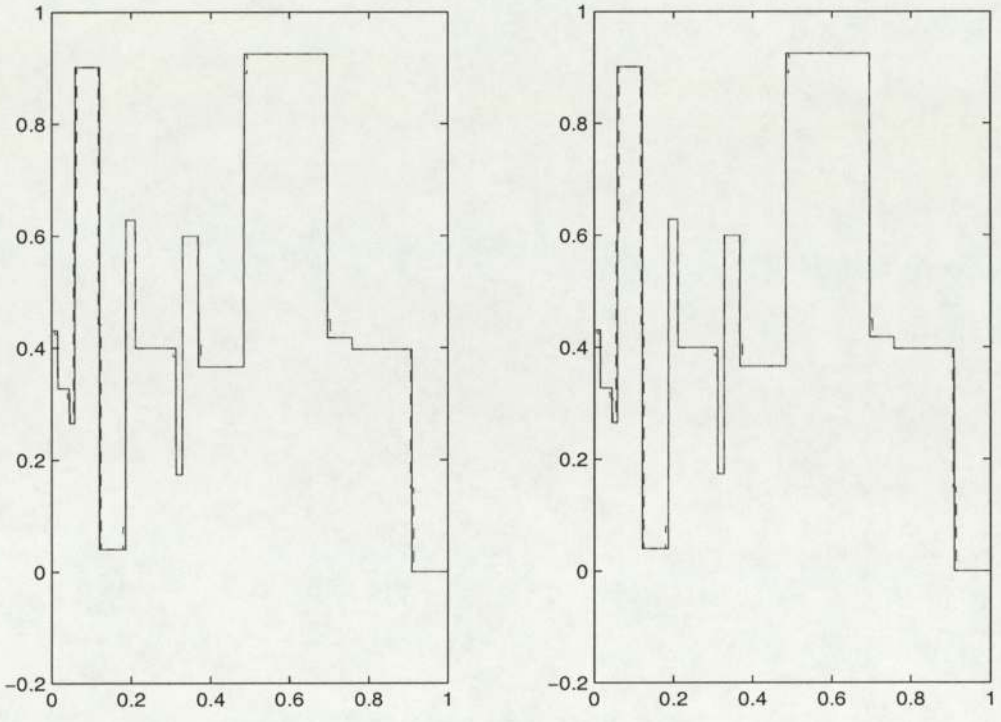

Figure 3.11: Approximation of the blocky signal using the dictionary B_3 with $b = 0.0625$. The approximations have been made using a different basis of B_3.

result for both dictionaries. Since the two bases create the same subspace, the approximation error, as shown in the first row of table 3.1, is almost the same. However after applying the BOOMP, we see that there is a big difference in the number of coefficients needed for representing the signal. In figure 3.13 we can see the two approximations after applying the BOOMP. The graph on the left corresponds to the approximation with the smaller error. In figure 3.14 we give a plot of the absolute value of the coefficients sorted in decreasing order. Stopping the procedure when we reach a certain value for the error, the approximations are almost the same. When we use the BOOMP algorithm to disregard coefficients up to an acceptable precision for the approximation, we see that the first basis needs only 15 coefficients (solid line), whereas with the second basis we have to use the double amount (30 coefficients, dashed line) and as it is clear from the table 3.1, the approximation is still not as good as the previous one.

	Basis 1	Basis 2
All the coefficients	$1.783 \cdot 10^{-13}$	$2.137 \cdot 10^{-13}$
After using BOOMP	0.0555	0.2943

Table 3.1: Representation errors, given by the norm of the difference between the signal and the calculated approximation.

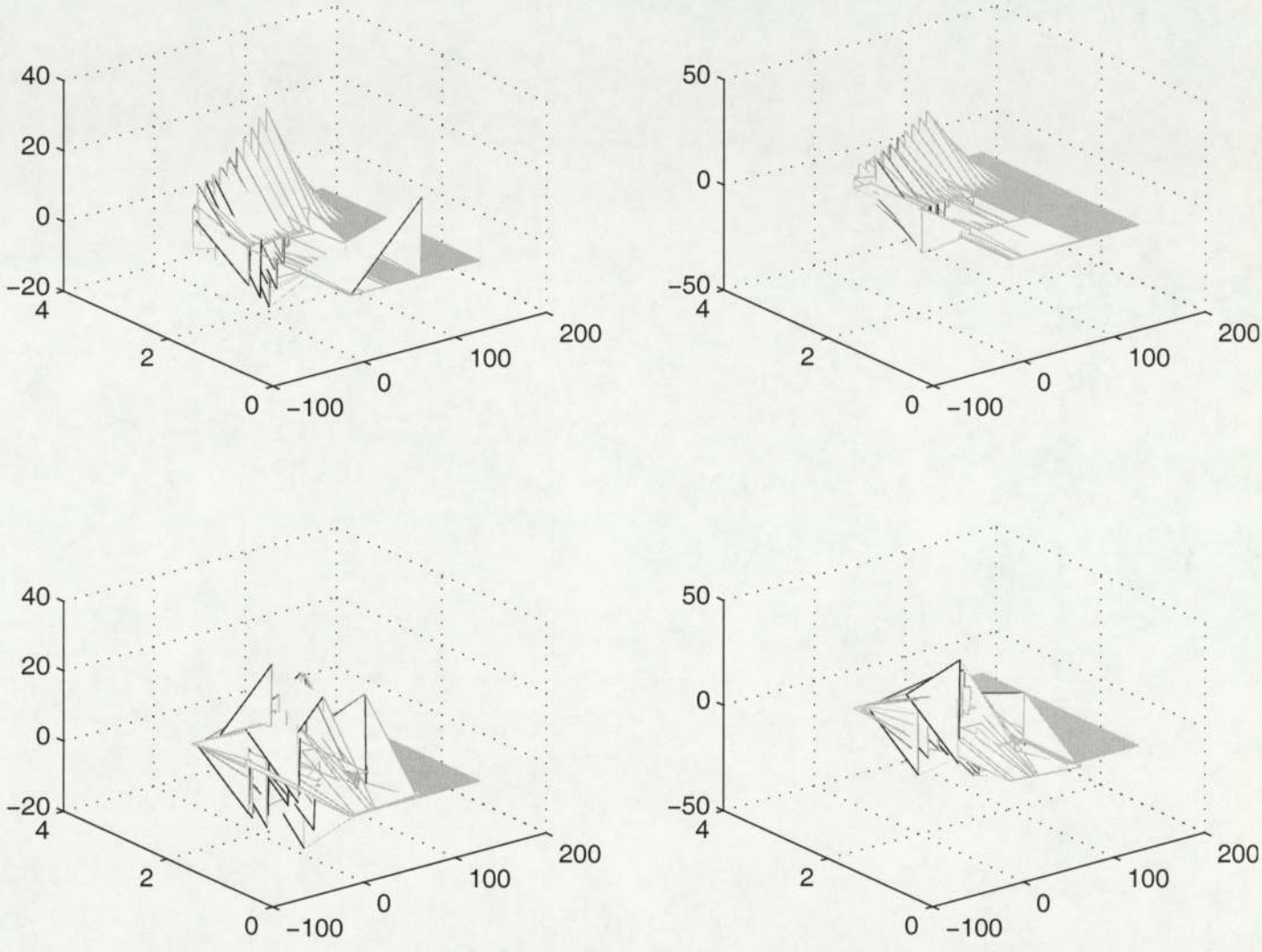

Figure 3.12: 3D plot of the coefficients for the approximation of the blocky function using two different bases. Before applying the BOOMP (upper row) and after (lower row).

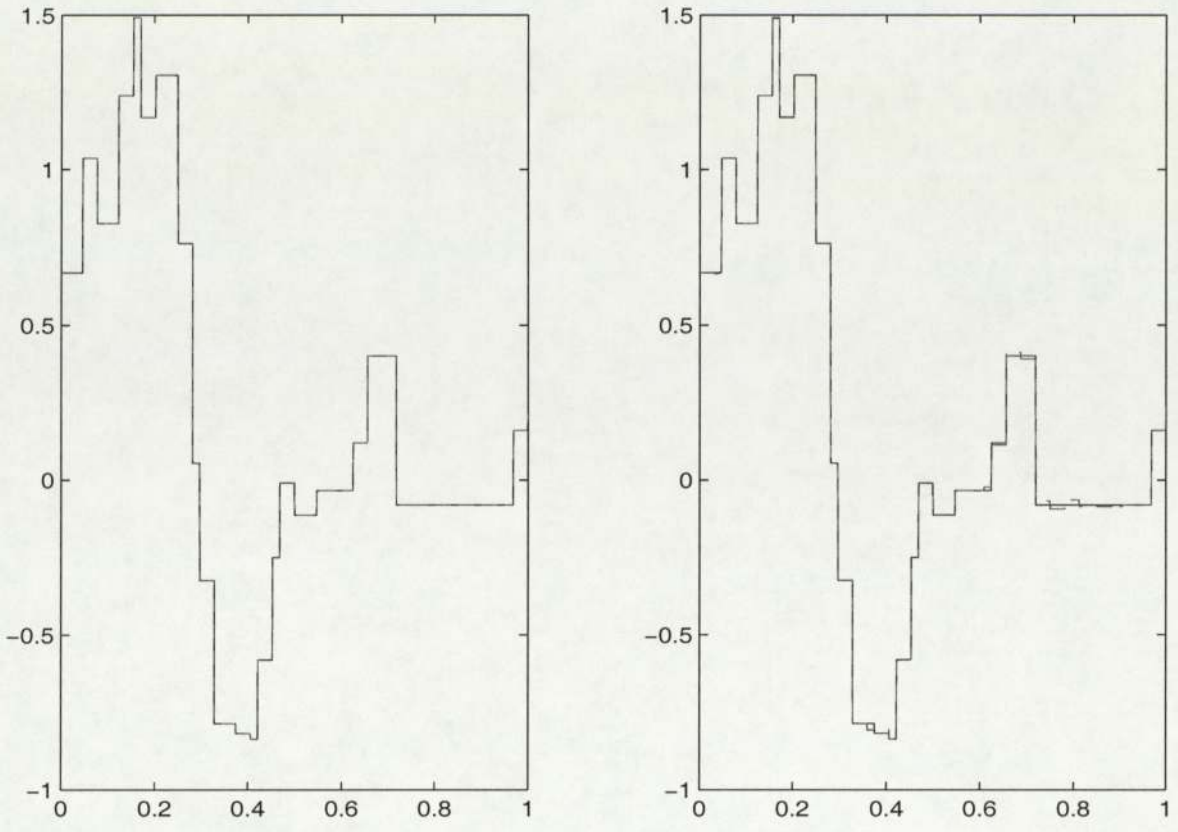

Figure 3.13: Approximation of a signal using two different bases that originate from the same dictionary.

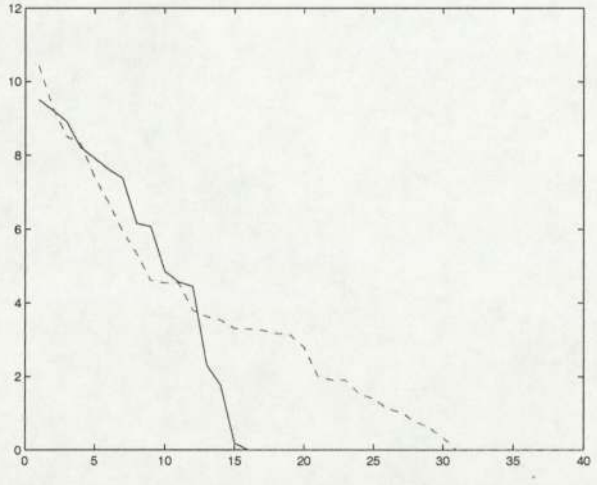

Figure 3.14: Graph of the absolute value of the coefficients sorted in decreasing order. The dashed line represents the coefficients needed for representing the signal of figure 3.12 (right graph). The solid line represents the coefficients needed for the approximation using a more suitable basis (left graph).

3.8.3 Conclusions

In this series of experiments we eliminated the redundancy from the non-orthogonal dictionary B_3 with translation parameter $b = 0.0625$. We obtained, thereby, different bases for the span of B_3. We have seen that the form of the corresponding biorthogonal functions may significantly change in each case. The experiments of this section illustrate that:

From the same dictionary we can pick different sets of linear independent atoms, i.e. different bases for the same subspace. Although all bases represent a given signal up to an identical precision, some bases may be more suitable than others, in terms of amount of coefficients and approximation error, for a specific signal.

Chapter 4

Experiments on signal representation

4.1 Introduction

In this chapter we illustrate the relevance of using appropriate dictionaries to represent different types of signals. To this end we will use two different dictionaries: The earlier introduced non-orthogonal Haar dictionary and the dictionary of mexican hat wavelets, to be introduced here. We will also consider the possibility of mixing these two dictionaries for representing signals of mixed features.

4.2 Mexican Hat dictionary

The *Mexican Hat* is a function obtained from the second derivative of the Gaussian $e^{-\frac{x^2}{2}}$. Its functional form is given by [5]

$$\psi(x) = C \cdot (1 - x^2) \cdot e^{-\frac{x^2}{2}}, \tag{4.1}$$

with $C = \frac{2}{\sqrt{3}}\pi^{-\frac{1}{4}}$ a normalisation constant. The shape of this function, as shown in 4.1, justifies its name.

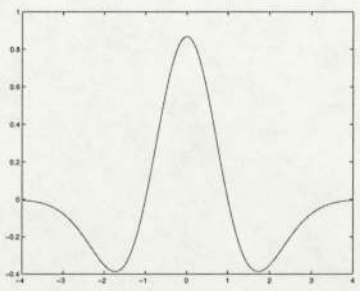

Figure 4.1: Mother wavelet for the mexican hat dictionary.

The mexican hat dictionary is composed of wavelet functions $\psi_{m,n}$ arising from the equation:

$$\psi_{m,n} = 2^{-\frac{m}{2}} \psi(a^m x - bn) \quad ; \quad m, n \in \mathbb{Z} \tag{4.2}$$

with, the mother wavelet $\psi_{0,0}(x) = \psi(x)$ as given in (4.1).

In figure 4.2 we plot some of the mexican hat wavelets for different scales. As shown in the graphs, changing the scale parameter m in (4.2) gives narrower or wider wavelets, while changing n gives translations of the same wavelet.

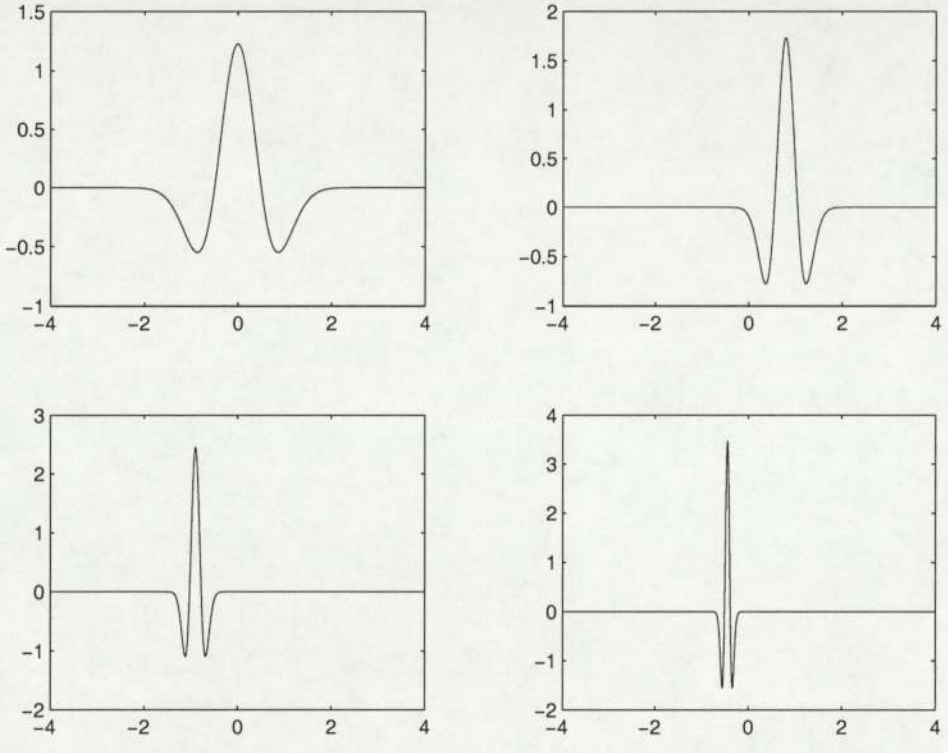

Figure 4.2: The mexican hat wavelets for different scale and translation parameters.

We denote the set of the mexican hat wavelets of scale j as $\Psi_j = \{\psi_{j,n}(x) \; ; \; n \in \mathbb{Z}\}$, and the set of mexican hat wavelets up to scale j as: $\Gamma_j = \Psi_0 \cup \Psi_1 \cup \ldots \cup \Psi_j$.

4.3 Comparing non-orthogonal dictionaries

4.3.1 Aims

Dictionaries of wavelets, like the Haar and the mexican hat dictionaries we have defined here, contain atoms which originate from the same mother wavelet. Thus, the shape of all the dictionary's elements is determined through that of the mother wavelet.

The aim of these experiments is to illustrate the results of choosing a suitable dictionary to approximate a given signal. These results are assessed in terms of number of coefficients in the representation and the corresponding residual error.

We will compare the mexican hat and the non-orthogonal Haar dictionaries to approximate two signals of different nature:

i) a modulated chirp and

ii) a blocky-type signal.

4.3.2 Procedure

First we approximate the modulated chirp signal in the interval [0,8]. We use the following two non-orthogonal dictionaries:

i) the mexican hat wavelets dictionary with translation parameter $b = 0.2$, and scaling parameter $a = 2$, denoted as Γ_4.

ii) The non-orthogonal Haar based dictionary from scales $m = -3$ to 2, denoted as $B_2 \cup B_{-3}/B_0$. The translation parameter of the functions is fixed as $b = 0.0625$.

In figure 4.3 (top left graph), we see that the mexican hat dictionary, using 75 coefficients, gives a high quality representation of the signal. In addition, by applying the BOOMP technique to this approximation, we obtain the top right graph with only 56 coefficients. Using the Haar based dictionary we obtain the approximation to the signal shown in the bottom left graph of figure 4.3 with 350 coefficients. After applying BOOMP to reduce the number of coefficients up to 130, the approximation is the one shown in the bottom right graph of the same figure. In both cases we stop the BOOMP algorithm for some level of approximation error.

Now we approximate a different type of signal, a blocky function in [0,8], with the same dictionaries as in the previous case. Figure 4.4 (top left graph) shows the approximation we obtain with the mexican hat dictionary and using 250 coefficients. Figure 4.4, (top right graph), shows the approximation after applying the BOOMP, to retain only 138 coefficients, for which the approximation is still acceptable.
With the non-orthogonal Haar dictionary we reach a satisfactory approximation of the signal, plotted with the dashed line in the bottom left graph of figure 4.4, using 102 coefficients. After applying BOOMP, retaining only 58 coefficients we still have a satisfactory approximation. This approximation is represented with the dashed line in the bottom right graph of the same figure.

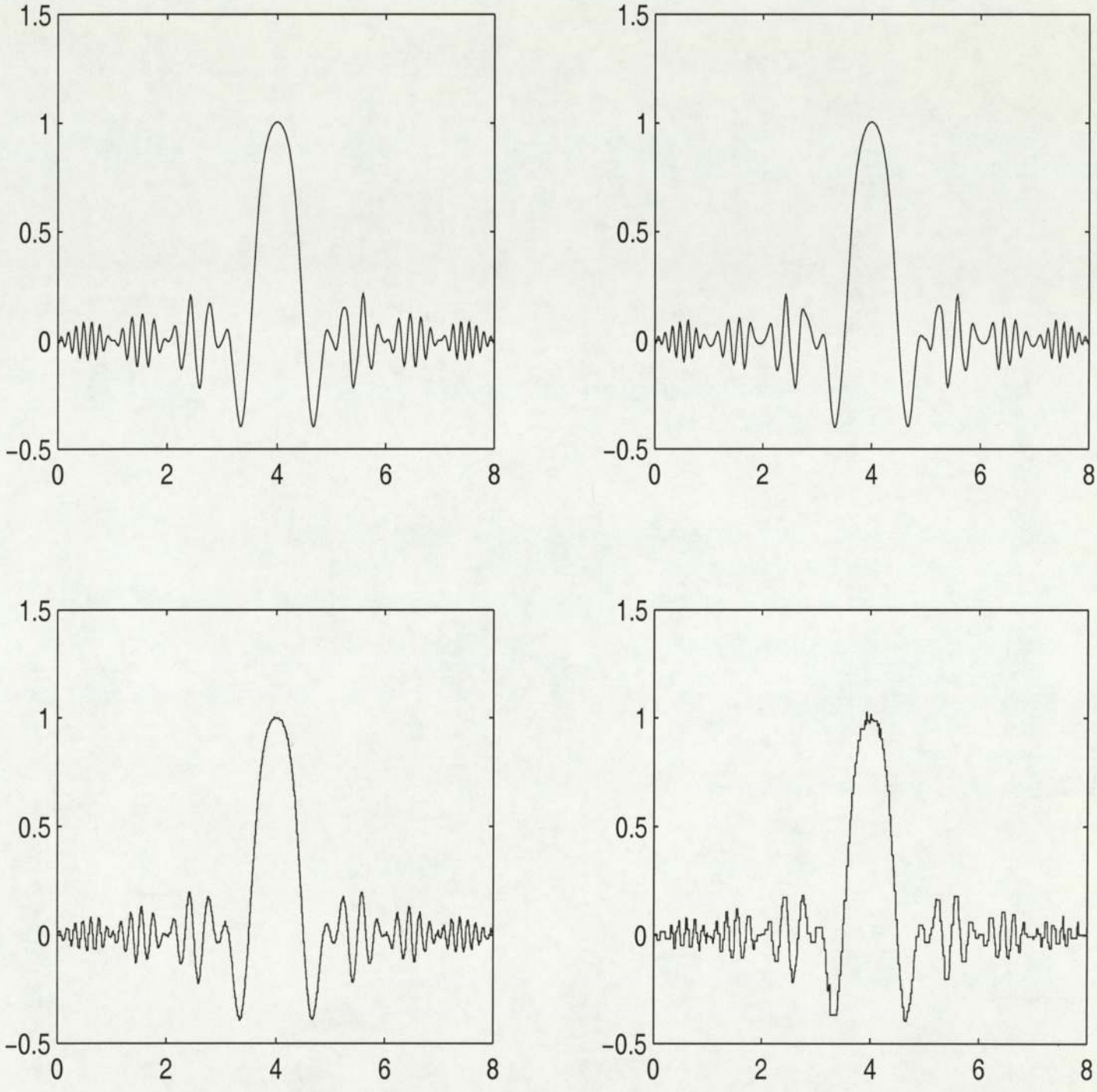

Figure 4.3: Top Left: Approximation of the chirp with the mexican hat, 75 coefficients. Top Right: Approximation using the mexican hat and BOOMP, 56 coefficients. Bottom Left: Using the non-orthogonal Haar, 350 coefficients. Bottom Right: Using the non-orthogonal Haar and BOOMP, 130 coefficients

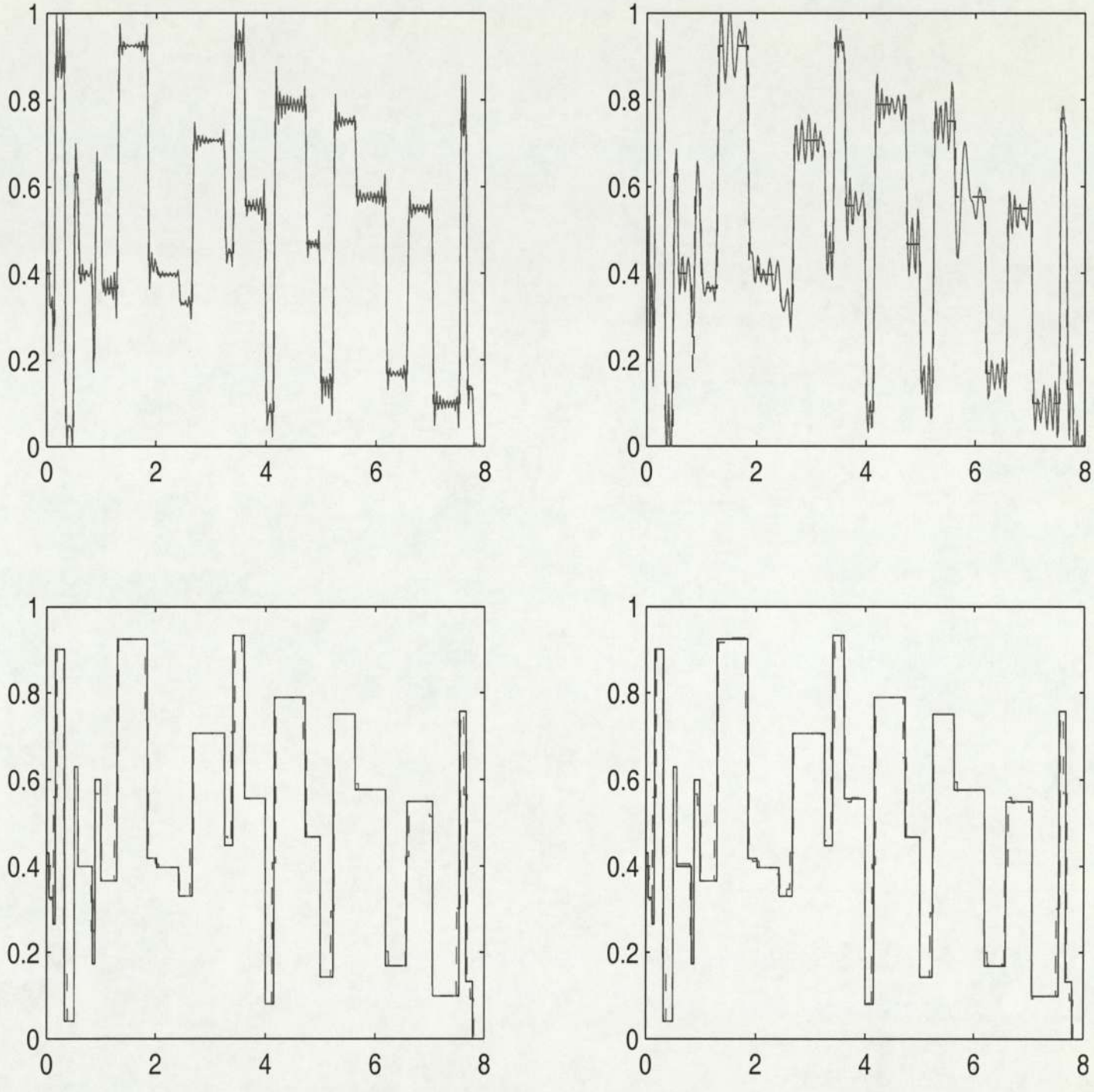

Figure 4.4: Top Left: Approximation of the blocky function using the mexican hat with 250 coefficients. Top Right: Using the mexican hat with BOOMP, 138 coefficients. Bottom Left: Using the non-orthogonal Haar, 102 coefficients. Bottom Right: Using the non-orthogonal Haar with BOOMP, 58 coefficients

4.3.3 Conclusions

In these experiments we have represented a modulated chirp and a blocky signal, both in the interval [0,8]. For the representation we have used the mexican hat and the non-orthogonal Haar dictionaries. We have compared the number of coefficients in the corresponding representations as well as the corresponding residual errors. As expected, the chirp signal is better approximated by the mexican hat dictionary and the blocky signal by the non-orthogonal Haar one. The results are of course not surprising but support the following intuitive remarks:

According to the type of signal we want to approximate, there are dictionaries that are more suitable than others. Signals with high regularity can be better approximated by dictionaries like the mexican hat, whereas signals with steps and discontinuities can be approximated better by dictionaries like the Haar. It is therefore very important to have the freedom to decide on the dictionary to be used in order to represent a signal of known properties.

4.4 Representation of a mixed signal

4.4.1 Aims

In the previous section we have seen that a dictionary may be more suitable than others to approximate a certain type of signal. We consider now the case in which the signal at hand is characterised by exhibiting distinct localised properties.

The aim of the experiments in this section is to analyse the possibility of using a mixed dictionary to approximate a signal of diverse features.

We will use the non-orthogonal Haar, the mexican hat dictionary and a dictionary composed by joining together the previous two.

4.4.2 Procedure

Let the signal we want to represent be the mixture of two localised blockies and a piece of low frequency chirp, plotted in the top left graph of figure 4.5. As discussed in the previous section the Haar dictionary is suitable for representing the blocky part of the signal and the mexican hat for the oscillating one. With the experiments of this section we will compare the approximations obtained when using each dictionary by itself, and the approximation obtained when using both dictionaries as one.

In figure 4.5 (top right graph), we plot the approximation of the signal obtained by using only the mexican hat dictionary and 224 coefficients (after applying BOOMP). In the blocky part of the signal we observe a phenomenon similar to the Gibbs effect for the Fourier basis. We should stress that, although by increasing the number of dictionary functions used in the approximation, the norm of the residual error does decrease (yet slowly), the Gibbs-type effect remains. This suggests that by using a dictionary of mexican hat wavelets we may not be able to have pointwise convergence to only piecewise continuous signals.

As it is seen in the bottom left graph of figure 4.5, by using a non-orthogonal Haar dictionary the Gibbs effect it is not present. The approximation plotted in the figure was obtained with 207 coefficients. However, in this case by using a larger number of coefficients (450) we can improve the approximation and make it visually identical to the exact signal.

We now consider a large set, containing the mexican hat and the non-orthogonal Haar dictionaries. Applying first the OOMP technique to the whole set and then the BOOMP on that approximation, we obtain the approximation of figure 4.5 (bottom right graph) with 104 coefficients. With respect to the previous approximations the number of coefficients has been reduced. Comparing the norm of the residual errors

given in table 4.1 we conclude that this approximation yields the smallest residual error by means of the smallest number of coefficients. However, the Gibbs effect in the blocky part of the signal persists. As in the case of the single mexican hat dictionary, although the norm of the residual error can be diminished by increasing the number of atoms in the representation, the Gibbs-type phenomenon remains. In other words: unfortunately the mixed dictionary has inherited from the mexican hat dictionary the non-pointwise convergence to only piecewise continuous signals.

	Coefficients	Error
mexican hat after BOOMP	224	0.0037
Non-orthogonal Haar after BOOMP	207	0.0035
mixed dictionary after BOOMP	104	0.0033

Table 4.1: For each approximation method we show the coefficients needed, and the corresponding residual error.

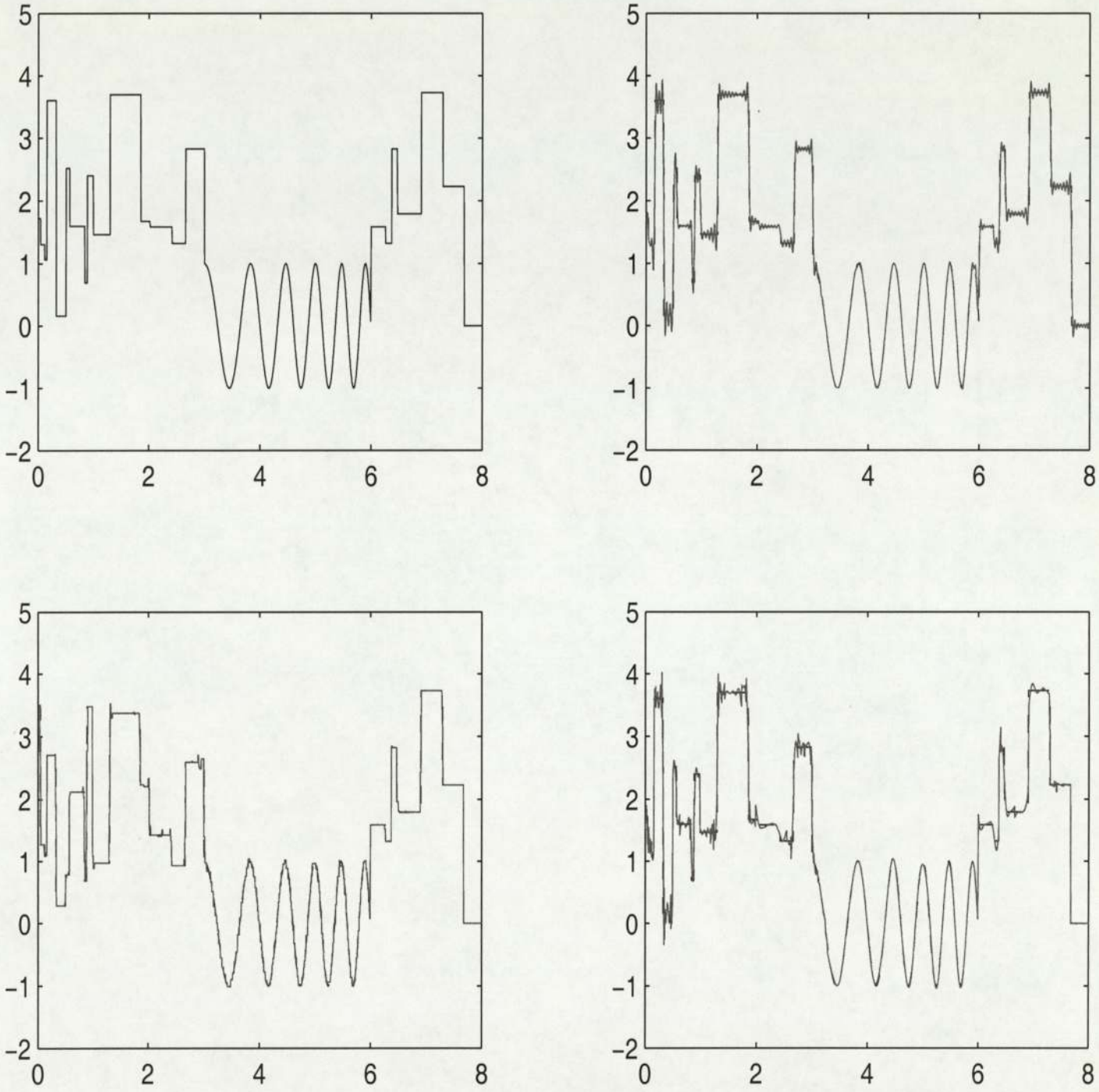

Figure 4.5: Top Left: The mixed signal: blockies with a low-frequency chirp. Top Right: Approximation using the mexican hat dictionary, 224 coefficients. Bottom Left: Approximation using the non-orthogonal Haar dictionary, 207 coefficients. Bottom Right: Approximation using the mixed dictionary, 104 coefficients.

4.4.3 Conclusions

These experiments presented here have given us some "first experience" on the use of mixed dictionaries to approximate signals of mixed features.

We understand that from this single experiment we cannot arrive to general conclusions. Nevertheless, we believe that we are in a position to safely assert that:

The use of mixed dictionaries for representing signals of mixed features seems to be an adequate procedure to achieve an economical representation. However, there is a convergence issue to be considered. It appears that, if one of the dictionaries does not yield pointwise convergence the mixture with other dictionaries does not guarantee pointwise convergence either.

It is clear that there is still much work to be done on this subject. Finding a smarter way of selecting atoms from mixed dictionaries would certainly be of great importance.

Chapter 5

Conclusions

5.1 General remarks

We divided the goal of the thesis in two main objectives. The first was to illustrate, by simple numerical examples, the advantage of choosing the waveforms to represent a signal by taking into account prior information about the signal properties. In general this entails having to resign the simplicity of dealing with orthogonal waveforms and face the nonlinear problem arising when using non-orthogonal ones. The techniques we have applied for this purpose can be classified in the line of adaptive matching pursuit. Such techniques operate by iteratively selecting waveforms (atoms) from a large, and in general redundant, dictionary of functions. In our experiments we have used the OOMP approach since, at least theoretically, it has been proved to improve upon earlier introduced MP techniques. We have illustrated this by a numerical example, since no simulation illustrating the theoretical improvement was available. We have also applied the recently introduced BOOMP, which allows for the reduction in the number of coefficients of a given non-orthogonal representation. This technique is easy to implement and our experiments show that its application may result in a significant gain as far as the economy of the representation is concerned. In addition, we have collected evidences suggesting that a recursive combination of OOMP and BOOMP techniques might result in a more economical representation. Unfortunately, due to time limitation we have not been able to look into the possibility of this combined strategy. We should leave the task as one of our main proposals of future work.

Our second objective was to be in a position to answer whether simply by relaxing the orthogonality condition one can obtain some gain in the signal representation problem. This has motivated the proposal of a generalisation of the well known orthogonal Haar dictionary. The new dictionary, that we have termed non-orthogonal Haar based dictionary, is generated simply by introducing a non-integer number in the translation parameter of the Haar wavelets. We found the results of this generalisation most

interesting. We can definitively conclude that in the finite dimensional case relaxing the orthogonality condition does not only introduce redundancy, but creates a larger subspace. Our experiments show that, for a fixed scale, by decreasing the translation parameters we obtain subspaces that can only be generated with orthogonal Haar functions of finer scale. The gain that can be achieved by using the new dictionary in problems of signal representation is clear. We are aware, however, that in order to use this dictionary for signal analysis, a new analysis scheme needs to be proposed. Looking into the possibility of devising such a scheme is our other main proposal of future work.

5.2 Future work proposals

- To combine forward/backward strategies in order to design new techniques for adaptive signal representation.
- To generalise other orthogonal wavelet families in order to see if the results holding by generalising the Haar dictionary are extensible to other cases.
- To look for a new wavelet analysis tool based on density of points in the translation parameter domain.
- To look for alternative criteria in choosing atoms from a mixed dictionary.

Appendix A

Projection Theorem

Let V be a closed subspace of a given Hilbert space $\mathcal{H}$. We denote by $V^{\perp}$ the set of elements in $\mathcal{H}$, which are orthogonal to V. The subspace $V^{\perp}$ is called the **orthogonal complement** of V, and it is also a Hilbert space.

Projection Theorem [17]: Let $f \in \mathcal{H}$, and V a closed subspace. Then every $f \in \mathcal{H}$ can be uniquely written:

$$f = g + h, \tag{A.1}$$

where $g \in V$ and $h \in V^{\perp}$.

From the above theorem we have the following proposition.

Proposition: The unique element $g \in V$ minimising the distance to an arbitrary element $f \in \mathcal{H}$, is obtained as: $g \equiv P_V f$, where with P_V we denote the orthogonal projector onto the subspace V.

Proof: Let $g \in \mathcal{H}$ be an arbitrary function in V. We write g as:

$$g = g - P_V f + P_V f.$$

Then the distance $\| f - g \|^2$ can be written as:

$$\| f - g \|^2 = \| f - g - P_V f + P_V f \|^2 . \tag{A.2}$$

Since $f - P_V f \in V^{\perp}$, from the projection theorem we have:

$$\| f - g \|^2 = \| f - P_V f \|^2 + \| P_V f - g \|^2 .$$

Thus, the distance $\| f - g \|^2$ is minimised if $g \equiv P_V f$.

Appendix B

Proposition

We give here the proof of property b) satisfied by functions arising from the recursive equations (2.26).

The following proposition was made in section 2.4.

Proposition: Let functions ψ_k and $\tilde{\psi}_{k+1}$ be given by equations (2.24) and (2.25), with $\psi_1 = \tilde{\psi}_1 = \alpha_{l_1}$. Let function $\beta_{l_1}^{(1)} = \alpha_{l_1}$ and the functions $\beta_{l_n}^{(k+1)}$ be defined as:

$$\beta_{l_n}^{(k+1)} = \beta_{l_n}^{(k)} - \frac{\tilde{\psi}_{k+1}}{\| \psi_{k+1} \|} \langle \alpha_{l_{k+1}}, \beta_{l_n}^{(k)} \rangle \quad ; \quad n = 1, \ldots, k \tag{B.1}$$

$$\beta_{l_{k+1}}^{(k+1)} = \frac{\tilde{\psi}_{k+1}}{\| \psi_{k+1} \|} \tag{B.2}$$

The above defined functions provide a representation of the orthogonal projection operator onto V_{k+1} as given by:

$$\hat{P}_{V_{k+1}} f = \sum_{n=1}^{k+1} \langle \beta_{l_n}^{(k+1)}, f \rangle \alpha_{l_n}. \tag{B.3}$$

Proof: The proof will be achieved by induction.

For k+1=1, we have: $\langle \beta_{l_1}^{(1)}, f \rangle \alpha_{l_1} = \langle \alpha_{l_1}, f \rangle \alpha_{l_1} = \hat{P}_{V_1} f$. Because the residual $R_1 = f - \langle \alpha_{l_1}, f \rangle \alpha_{l_1}$ is orthogonal to α_{l_1}, i.e. $\langle \alpha_{l_1}, R_1 \rangle = 0$.

Assuming that (B.3) holds for k, we will show that it is also true for $k+1$. Using

(B.1), (B.2) we have:

$$\sum_{n=1}^{k+1}\langle\beta_{l_n}^{(k+1)}, f\rangle\alpha_{l_n} = \sum_{n=1}^{k}\langle\beta_{l_n}^{(k+1)}, f\rangle\alpha_{l_n} + \langle\frac{\tilde{\psi}_{k+1}}{\| \psi_{k+1} \|}, f\rangle\alpha_{l_{k+1}}$$
$$= \sum_{n=1}^{k}\langle\beta_{l_n}^{(k)}, f\rangle\alpha_{l_n} - \sum_{n=1}^{k}\langle\beta_{l_n}^{(k)}, \alpha_{l_{k+1}}\rangle\langle\frac{\tilde{\psi}_{k+1}}{\| \psi_{k+1} \|}, f\rangle\alpha_{l_n} + \langle\frac{\tilde{\psi}_{k+1}}{\| \psi_{k+1} \|}, f\rangle\alpha_{l_{k+1}}. \qquad \text{(B.4)}$$

By hypothesis

$$\sum_{n=1}^{k}\langle\beta_{l_n}^{(k)}, f\rangle\alpha_{l_n} = \hat{P}_{V_k} f$$
$$\text{and} \quad \sum_{n=1}^{k}\alpha_{l_n}\langle\beta_{l_n}^{(k)}, \alpha_{l_{k+1}}\rangle = \hat{P}_{V_k}\alpha_{l_{k+1}},$$

so that from (B.4) we have:

$$\sum_{n=1}^{k+1}\langle\beta_{l_n}^{(k+1)}, f\rangle\alpha_{l_n} = \hat{P}_{V_k} f - \hat{P}_{V_k}\alpha_{l_{k+1}}\langle\frac{\tilde{\psi}_{k+1}}{\| \psi_{k+1} \|}, f\rangle + \langle\frac{\tilde{\psi}_{k+1}}{\| \psi_{k+1} \|}, f\rangle\alpha_{l_{k+1}}$$
$$= \hat{P}_{V_k} f + (\alpha_{l_{k+1}} - \hat{P}_{V_k}\alpha_{l_{k+1}})\langle\frac{\tilde{\psi}_{k+1}}{\| \psi_{k+1} \|}, f\rangle$$
$$= \hat{P}_{V_k} f + \hat{P}_{W_{k+1}} f = \hat{P}_{V_{k+1}} f. \qquad \text{(B.5)}$$

Thus, it is proved that (B.3) holds for all k.

Appendix C

Theorem I

We give here the proof of Theorem I, (section 2.4).

Theorem I: The dictionary atom $\alpha_{l_{k+1}}$ that at iteration $k+1$ minimises the norm of the residual $\tilde{R}_{k+1}$ is the one that gives the maximal value for the expression:

$$e_n = \frac{|\langle \alpha_n, \tilde{R}_k \rangle|^2}{1 - \langle \alpha_n, \hat{P}_{V_k} \alpha_n \rangle} = \frac{|\langle \psi_n, f \rangle|^2}{\| \psi_n \|^2} \quad ; \quad \| \psi_n \| \neq 0, \tag{C.1}$$

where for each n the corresponding function ψ_n is obtained by:

$$\psi_n = \alpha_n - \hat{P}_{V_k} \alpha_n. \tag{C.2}$$

Proof: Since the $k+1$ order residual is given by: $\tilde{R}_{k+1} = f - \hat{P}_{V_{k+1}} f$, we have: $\| \tilde{R}_{k+1} \|^2 = \| f \|^2 - \langle \hat{P}_{V_{k+1}} f, f \rangle$. Thus, in order to minimise $\| \tilde{R}_{k+1} \|$ we have to maximise the expression: $\langle \hat{P}_{V_{k+1}} f, f \rangle$. Using the equation:

$$\sum_{n=1}^{k+1} \langle \beta_{l_n}^{(k+1)}, f \rangle \alpha_{l_n} = \hat{P}_{V_k} f + \hat{P}_{W_{k+1}} f = \hat{P}_{V_{k+1}} f,$$

we have:

$$\langle \hat{P}_{V_{k+1}} f, f \rangle = \langle \hat{P}_{V_k} f, f \rangle + \frac{1}{\| \psi_{k+1} \|^2} |\langle \alpha_{l_{k+1}}, f \rangle - \langle \hat{P}_{V_k} \alpha_{l_{k+1}}, f \rangle|^2. \tag{C.3}$$

So that to maximise the expression (C.3), considering that $\langle \hat{P}_{V_k} f, f \rangle$ is fixed, we have to maximise:

$$\frac{|\langle \alpha_{l_{k+1}}, f \rangle - \langle \hat{P}_{V_k} \alpha_{l_{k+1}}, f \rangle|^2}{\| \psi_{k+1} \|^2} = \frac{|\langle \alpha_{l_{k+1}}, f - \hat{P}_{V_k} f \rangle|^2}{1 - \langle \alpha_{l_{k+1}}, \hat{P}_{V_k} \alpha_{l_{k+1}} \rangle} = \frac{|\langle \alpha_{l_{k+1}}, \tilde{R}_k \rangle|^2}{1 - \langle \alpha_{l_{k+1}}, \hat{P}_{V_k} \alpha_{k+1} \rangle}. \tag{C.4}$$

By noticing that the left hand side of the equation is equal to $\frac{|\langle \psi_{l_{k+1}}, f \rangle|^2}{\| \psi_{l_{k+1}} \|^2}$ the proof is concluded.

Appendix D

Theorem II

We give here the proof of Theorem II (section 2.6):

Theorem II: Let $\tilde{R}_j$ be the residual resulting by disregarding a coefficient $c_j^{(k)}$ as passing from approximation f_{V_k} to $f_{V_{k/\alpha_j}}$. In order to minimise the norm of the residual $\tilde{R}_j$ such coefficient is to be chosen as the one yielding a minimum value of the quantity:

$$\frac{|c_j^{(k)}|^2}{\| \beta_j^{(k)} \|^2}. \tag{D.1}$$

Proof: Since $\tilde{R}_j = f_{V_k} - f_{V_{k/\alpha_j}} = \sum_{n=1}^{k} c_n^{(k)} \alpha_n - \sum_{\substack{n=1 \\ n \neq j}}^{k} c_n^{(k/j)} \alpha_j$, by using equation (2.37), we have:

$$\begin{aligned} \tilde{R}_j &= \sum_{n=1}^{k} c_n^{(k)} \alpha_n - \sum_{\substack{n=1 \\ n \neq j}}^{k} c_n^{(k)} \alpha_n + \sum_{\substack{n=1 \\ n \neq j}}^{k} c_j^{(k)} \alpha_n \frac{\langle \beta_n^{(k)}, \beta_j^{(k)} \rangle}{\| \beta_j^{(k)} \|^2} \\ &= c_j^{(k)} \alpha_j + c_j^{(k)} \sum_{n=1}^{k} \alpha_n \frac{\langle \beta_n^{(k)}, \beta_j^{(k)} \rangle}{\| \beta_j^{(k)} \|^2} - c_j^{(k)} \alpha_j. \end{aligned} \tag{D.2}$$

Since $\sum_{n=1}^{k} \alpha_n \langle \beta_n^{(k)}, \beta_j^{(k)} \rangle = \beta_j^{(k)}$ it follows from the last equation, that:

$$\tilde{R}_j = \frac{c_j^{(k)} \beta_j^{(k)}}{\| \beta_j^{(k)} \|^2}.$$

Hence, the coefficient minimising the expression (D.1) yields the minimal value of $\| \tilde{R}_j \|^2$.

List of Figures

List of Tables

Bibliography

[1] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM , Journal on Scientific Computing*, 20:33–61, 1998.

[2] R. R. Coifman and M. V. Wickerhauser. Entropy-based algorithms for best-basis selection. *IEEE Transactions on Information Theory*, 38:713–718, 1992.

[3] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Commun. on Pure and Applied Mathematics*, 41:909–996, 1988.

[4] I. Daubechies. The wavelets transform,time frequency localisation and signal analysis. *IEEE Transactions on Information Theory*, 27:961–1005, 1990.

[5] I. Daubechies. *Ten Lectures on Wavelets.* SIAM, 1992.

[6] G. M. Davis, S. Mallat, and M. Avellaneda. Adaptive greedy approximations. *J. of Conts. Approx.*, 13:57–98, 1997.

[7] S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, 1998.

[8] S. Mallat and Z. Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41 (12):3397–3415, 1993.

[9] Y. Meyer. *Wavelets, Algorithms and Applications.* SIAM, 1993.

[10] A. Papoulis. *Signal Analysis.* McGraw-Hill, New York, 1977.

[11] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. Orthogonal matching pursuits: recursive function approximation with applications to wavelet decomposition. *Proceedings of the 27th Asilomar Conference in Signals, Systems and Computers*, 1993.

[12] L. Rebollo-Neira. Backward adaptive biorthogonalization. *math-ph/0211066*, 2002.

[13] L. Rebollo-Neira. Backward optimised orthogonal matching pursuit approach. *pre print*, 2002.

[14] L. Rebollo-Neira. Recursive biorthogonalisation approach and orthogonal projectors. *math-ph/0209026*, 2002.

[15] L. Rebollo-Neira. Dictionary redundancy elimination. *IEE Proceedings - Vision, Image and Signal Processing*, 2003.

[16] L. Rebollo-Neira and D. Lowe. Optimised orthogonal matching pursuit approach. *IEEE Signal Processing Letters*, 9,4:137,140, 2002.

[17] M. Reed and B. Simon. *Functional Analysis.* Academic Press, 1980.

[18] B. Vidacovic. *Statistical Modelling by Wavelets.* John Wiley and Sons, Inc., 1999.

[19] D. F. Walnut. *An Introduction to Wavelet Analysis.* Birkhäuser, 2002.

[20] D. S. Watkins. *Fundamentals of Matrix Computations,Second Edition.* John Wiley and Sons, Inc., 2002.

[21] R. M. Young. *An Introduction to non-harmonic Fourier Series.* Academic Press, New York, 1980.